

RTE

Revista Tecnológica
ESPOL

Vol. 35, N° 2 (2023)

ISSN 0257-1749

e-ISSN 1390-3659

DOI 10.37815/rte

LAS TIC Y SU IMPACTO EN LA SOCIEDAD



TRACK TÉCNICO

Edición Especial:
TICEC2023



Guayaquil, Ecuador
Vol. 35, Núm. 2 (2023)
Octubre, 15

DOI: 10.37815/rte.v35n2



Esta obra está bajo una licencia internacional
Creative Commons Atribución-NoComercial 4.0.

Autoridades Institucionales (Institutional Authorities)

Rectora

Cecilia Alexandra Paredes Verduga, Ph.D.
 Escuela Superior Politécnica del Litoral, Ecuador

Vicerrectora de Docencia

Paola Romero, Ph.D.
 Escuela Superior Politécnica del Litoral, Ecuador

Vicerrectorado de Investigación, Desarrollo e Innovación (I+D+i)

Carlos Teodoro Monsalve Arteaga, Ph.D.
 Escuela Superior Politécnica del Litoral, Ecuador

Consejo de Editores (Editorial Board)

Directora General Editorial (Editor in Chief)

Nayeth I. Solórzano Alcivar, Ph.D.
 Escuela Superior Politécnica del Litoral, Ecuador
 e-mail: nsolorza@espol.edu.ec
 ORCID: 0000-0002-5642-334X

Directora Ejecutiva Editorial -Coeditor (Executive Director)

Katherine Anabelle Salvador Cisneros, Ph.D.
 Escuela Superior Politécnica del Litoral, Ecuador
 e-mail: ksalvado@espol.edu.ec
 ORCID: 0000-0003-0414-3597

Editores Asociados (Associate Editors)

Fausto Enrique Jácome López, M.Sc.
 Escuela Superior Politécnica del Litoral, Ecuador
 e-mail: fjacome@espol.edu.ec
 ORCID: 0000-0001-6808-7294

Santiago Berrezueta-Guzman, Ph.D.
 Technical University of Munich - TUM, Germany
 Corporación Ecuatoriana para el Desarrollo de la Investigación y la Academia - CEDIA, Ecuador
 e-mail: santiago.berrezueta@cedia.org.ec
 ORCID: 0000-0001-5559-2056

Editores/as Invitados (Invited Editors)

Irma Abrigo Córdova, Ph.D.
 Universidad Nacional de Loja, Ecuador
 e-mail: irmabrigocordova2016@gmail.com
 ORCID: 0000-0002-1537-6313

Patrick Taran, Prof.
 Global Migration Policy Associates, GMPA,
 Switzerland
 e-mail: taran@globalmigrationpolicy.org
 ORCID: 0000-0002-9442-5725

Editores/as de Sección o Temático (Section Editors)

Marcelo Rafael Báez Meza, Ph.D.
 Escuela Superior Politécnica del Litoral, Ecuador
 e-mail: mbaez@espol.edu.ec
 ORCID: 0000-0002-8563-6985

Miguel Eduardo Yapur Aua, M.Sc.
 Escuela Superior Politécnica del Litoral, Ecuador
 e-mail: myapur@espol.edu.ec
 ORCID: 0000-0001-6627-7908

Silvana del Pilar Gamboa Benitez, Ph.D.
 Escuela Politécnica Nacional, Ecuador
 e-mail: silvana.gamboa@epn.edu.ec
 ORCID: 0000-0001-5951-1970

Sylvia Novillo Villegas, Ph.D.
 Universidad de las Américas, Ecuador
 e-mail: sylvia.novillo@udla.edu.ec
 ORCID: 0000-0002-7659-6846

Ramón Espinel Martínez, Ph.D.
 Centro de investigaciones rurales, CIR
 Escuela Superior Politécnica del Litoral, Ecuador
 e-mail: respinel@espol.edu.ec
 ORCID: 0000-0002-4998-2063

Marco Antonio González Escudero, Dr.
 Universidad Central del Ecuador, Ecuador
 e-mail: magonzalez@uce.edu.ec
 ORCID: 0000-0003-4415-6579

Fernando Sandoya Sánchez, Ph.D.
 Universidad de Guayaquil, Ecuador
 e-mail: fernando.sandoyas@ug.edu.ec
 ORCID: 0000-0002-0011-4003

Cristian Díaz Álvarez, Ph.D.
 Universidad del Meta, Colombia
 e-mail: cristian.diaz@unimeta.edu.co
 ORCID: 0000-0002-4212-4947

Eduardo Jordan Pérez, Ph.D.
 Griffith University, Australia
 e-mail: e.jordanperez@griffith.edu.au
 ORCID: 0000-0001-5299-8501

Wellington Enrique Wonsang Ureta, BS.c.
San Diego State University Associates in Business Accounting, Estados Unidos
e-mail: wonsangwellington@gmail.com
ORCID: 0000-0003-3936-9705

Rosa Aurora Espinoza Toalombo, Ph.D.
Universidad Estatal de Milagro, Ecuador
e-mail: respinozat@unemi.edu.ec
ORCID: 0000-0001-6527-1870

Editores/as de Copia - Español (Copy Editors - Spanish)

Diana Llamel Rodríguez Arteaga, M.Sc.
Escuela Superior Politécnica del Litoral, Ecuador
e-mail: dirodrig@espol.edu.ec
ORCID: 0000-0002-3715-181X

Andrea Victoria Carranza García, M.Sc.
Escuela Superior Politécnica del Litoral, Ecuador
e-mail: avcarran@espol.edu.ec
ORCID: 0000-0003-1762-4525

Laura Amelia Malache Silva, Lcda.
CEDIA, Ecuador
e-mail: laura.malache@cedia.org.ec
ORCID: 0009-0004-2846-8710

Editores/as de Copia - Inglés (Copy Editors - English)

Lidia Govea de Bustamante, Ph.D.
Escuela Superior Politécnica del Litoral, Ecuador
e-mail: bgovea@espol.edu.ec
ORCID: 0000-0002-6770-1261

Maite Velez Chamaidan, Lic.
Massachusetts Bay Community College,
Estados Unidos
e-mail: m_velez3@massbay.edu
ORCID: 0009-0008-8633-818X

Consejo Científico (Advisory Board)

Carlos Teodoro Monsalve Arteaga, Ph.D.
Escuela Superior Politécnica del Litoral, Ecuador
e-mail: monsalve@espol.edu.ec
ORCID: 0000-0002-9428-4360

Boris Xavier Vintimilla Burgos, Ph.D.
Escuela Superior Politécnica del Litoral, Ecuador
e-mail: bvintim@fiec.espol.edu.ec
ORCID: 0000-0001-8904-0209

Carlos Alberto Ramos Galarza, Ph.D.
Pontificia Universidad Católica del Ecuador
e-mail: ps_carlosramos@hotmail.com
ORCID: 0000-0001-5614-1994

Diana Carolina Coello Fiallos, Ph.D.
Universidad Técnica de Ambato, Ecuador
e-mail: dc.coello@uta.edu.ec
ORCID: 0000-0003-3628-5004

Diego Fernando Avila Pesantez, Ph.D.
Escuela Superior Politécnica de Chimborazo,
Ecuador
e-mail: davila@epoch.edu.ec
ORCID: 0000-0001-8394-5621

Elena María Fernández González, Ph.D.
Universidad de Oviedo, España
e-mail: fernandezelena@uniovi.es
ORCID: 0000-0002-5683-0674

Elva Patricia Reyes Díaz, Ph.D.
Universidad Autónoma de Coahuila, México
e-mail: elva.reyes@uadec.edu.mx
ORCID: 0000-0002-8854-0946

Ignacio Bladimir Cerón Guerra, Ph.D.
Escuela Politecnica Nacional, Ecuador
e-mail: ignacio.ceron@epn.edu.ec
ORCID: 0000-0002-6373-9772

Isidro Fierro Ulloa, Ph.D.
Universidad de Especialidades Espíritu Santo,
Ecuador
e-mail: isfierro@uees.edu.ec
ORCID: 0000-0002-9000-6945

José Enrique Martínez-Pérez, Ph.D.
Instituto Universitario de Tecnología del Estado
Bolívar, Venezuela
e-mail: josenriquemartinez@gmail.com
ORCID: 0000-0003-3611-639

Jorge Antonio Córdova Morán, Ph.D.
Universidad Estatal de Milagro, Ecuador
e-mail: joranto205@gmail.com
ORCID: 0000-0002-2730-4755

Juan José Oliveira Maurera, Ph.D.
Universidad de Oriente, Venezuela
e-mail: juanjoseoli@hotmail.com
ORCID: 0000-0002-8474-5872

Juan Humberto Avellaneda Cevallos, Ph.D.
Universidad Técnica Estatal de Quevedo,
Ecuador
e-mail: juan_avellaneda@yahoo.com
ORCID: 0000-0002-1805-4803

Luke Houghton, Ph.D.
Griffith University, Australia
e-mail: l.houghton@griffith.edu.au
ORCID: 0000-0002-7461-6591

Manuel Raúl Peláez Samaniego, Ph.D.
 Universidad de Cuenca, Ecuador
 e-mail: manuel.pelaez@ucuenca.edu.ec
 ORCID: 0000-0002-7618-9474

María Del Pilar Cornejo Rodríguez, Ph.D.
 Escuela Superior Politécnica del Litoral, Ecuador
 e-mail: pcornejo@espol.edu.ec
 ORCID: 0000-0003-4198-2904

María Genoveva Espinoza-Santeli, Ph.D.
 Universidad Andina Simón Bolívar, Ecuador
 e-mail: genoveva.espinoza@uasb.edu.ec
 ORCID: 0000-0002-2347-9507

Olga Kadysheva, Ph.D.
 Global Migration Policy Associates, Switzerland
 e-mail: okadysheva@globalmigrationpolicy.org
 ORCID: 0000-0002-4810-8964

Patricia A. Boeri, Ph.D.
 Universidad Nacional de Río Negro, Argentina
 e-mail: pboeri@unrn.edu.ar
 ORCID: 0000-0003-3262-8030

Paúl Cesar Carrión Mero, Ph.D.
 Escuela Superior Politécnica del Litoral, Ecuador
 e-mail: pcarrión@espol.edu.ec
 ORCID: 0000-0002-9747-7547

Raquel Miroslava Tinoco Egas, Ph.D.
 Universidad Técnica de Machala, Ecuador
 e-mail: raqueltinocoegas@hotmail.com
 ORCID: 0000-0001-5979-8123

Stanislaus Albert-Georg Sonnenholzner, Ph.D.
 Escuela Superior Politécnica del Litoral, Ecuador
 e-mail: ssonnen@cenaim.espol.edu.ec
 ORCID: 0000-0002-5723-3994

Consejo Internacional de Revisores del Volumen (International Reviewers Board)

Ana-Lucía Pérez-Suasnavas
 Universidad Central del Ecuador, Ecuador
 e-mail: alperezs@uce.edu.ec
 ORCID: 0000-0002-1700-9672

Efraín Rodrigo Fonseca Carrera, Dr
 Universidad de las Fuerzas Armadas, Ecuador
 e-mail: erfonseca@espe.edu.ec
 ORCID: 0000-0001-7129-9335

Fernando Romero
 Universidad de Cuenca
 e-mail: feroqa1656@hotmail.com
 ORCID: 0000-0002-6836-0251

Gandy Rene López Fuertes, Ph.D.
 Ministerio de Inclusión Económica y Social,
 Ecuador
 e-mail: gandy.lopez@rocketmail.com
 ORCID: 0000-0001-9101-8932

Ivan de la Cruz
 Universidad Politécnica de Madrid, España
 e-mail: ivan.delacruz84@outlook.es
 ORCID: 0000-0001-8413-2972

Jimmy Javier Cordova Villagomez, M.Sc.
 Escuela Superior Politécnica del Litoral, Ecuador
 e-mail: jjcordov@espol.edu.ec
 ORCID: 0000-0002-4082-4923

Jonnathan Guzman
 Universidad Técnica de Múnich, Alemania
 e-mail: ge45baj@mytum.de
 ORCID: 0009-0006-7721-8018

Jorge Herrera
 Universidad Laica Eloy Alfaro de Manabí,
 Ecuador
 e-mail: jorge.herrera@uleam.edu.ec
 ORCID: 0000-0002-8673-0236

Jorge Luis Zambrano-Martinez, Dr.
 Universidad Del Azuay, Ecuador;
 Universitat Poliècnica de València, España
 e-mail: jorge.zambrano@uazuay.edu.ec
 ORCID: 0000-0002-5339-7860

Juan Manuel Peralta Jaramillo, Dr.
 Escuela Superior Politécnica del Litoral, Ecuador
 e-mail: jperal@espol.edu.ec
 ORCID: 0000-0001-9812-2281

María Genoveva Espinoza-Santeli, Ph.D.
 Universidad Andina Simón Bolívar, Ecuador
 e-mail: genoveva.espinoza@uasb.edu.ec
 ORCID: 0000-0002-2347-9507

Maria Luisa Ruiz
 Universidad Politécnica de Madrid, España
 e-mail: m.luisa.ruiz.m29@gmail.com
 ORCID: 0000-0001-7238-1162

Melissa Montalvo
 ScreenPoint Medical, Los Países Bajos
 e-mail: melissa.montalvo@screenpointmed.com
 ORCID: 0000-0002-9539-3243

Miguel Ángel Gutiérrez Soto, Mgs.
 Universidad Adventista de Chile, Chile
 e-mail: miguelagutierrez@unach.cl
 ORCID: 0000-0003-0048-9208

Paula Monticelli
 Universidad Politécnica de Madrid, España
 e-mail: paula.monticelli79@outlook.com
 ORCID: 0000-0003-2338-5639

Pedro Fabricio Zanzzi Diaz, Ph.D.
Escuela Superior Politécnica del Litoral, Ecuador
e-mail: pzanzzi@espol.edu.ec
ORCID: 0000-0002-1396-6854

Rosa Aurora Espinoza Toalombo, Ph.D.
Universidad Estatal de Milagro, Ecuador
e-mail: respinozat@unemi.edu.ec
ORCID: 0000-0001-6527-1870

Silvia Mariela Mendez Prado, M.Sc.
Escuela Superior Politécnica del Litoral, Ecuador
e-mail: smendez@espol.edu.ec
ORCID: 0000-0003-2978-2424

Gestión de Comunicación, Publicación y Técnica

Difusión y Comunicación Gerencia de Comunicación Social y Asuntos Públicos, ESPOL

José Arturo Maldonado Moncayo, Ing.
Escuela Superior Politécnica del Litoral, Ecuador

Community Manager

Patricia Alexandra Caballero Roldan, M.Sc
Escuela Superior Politécnica del Litoral, Ecuador

Customer Service

Daniela Isabela Villón Martínez
Escuela Superior Politécnica del Litoral, Ecuador

Traducciones en idioma extranjero

English

Joyce Nan, BA.
Griffith University, Australia

Portada

Daniela Chuquiguanga Morales
CEDIA, Ecuador

Director de Arte

Paúl Arévalo García
CEDIA, Ecuador

Diseño y Diagramación

Kleber José Avelino Mosquera
Escuela Superior Politécnica del Litoral, Ecuador

Informática

Contacto Soporte Técnico
Kleber José Avelino Mosquera
rte@espol.edu.ec
Escuela Superior Politécnica del Litoral, Ecuador

Indexadores / Directorios



e-revist@s

REDIB | Red Iberoamericana
de Innovación y Conocimiento Científico



Google
Scholar

publons

ROAD



INDEX
COPERNICUS

MIAR

Identificadores

ORCID
Connecting Research and Researchers



Sobre la revista

La **Revista Tecnológica ESPOL - RTE** es una publicación digital, a la que se puede acceder de forma abierta y gratuita desde su sitio web <http://www.rte.espol.edu.ec>. Se encuentra registrada en varios indexadores y directorios de bases de datos y cuenta con un proceso de evaluación continua de nuevas formas de indexación que permiten dar mayor visibilidad a los artículos publicados y a sus autores.

RTE presenta llamadas programadas para envío de artículos a través de la plataforma Open Journal System, garantizando como mínimo la publicación de dos volúmenes al año. Sin embargo, cada año se producen y se garantiza volúmenes adicionales sobre una temática específica (como un Monográfico o Dossier) o ediciones especiales relacionadas a un evento científico o tecnológico, que recogen artículos o ponencias revisadas y aceptadas por pares ciegos antes de su publicación.

About

The **Revista Tecnológica ESPOL - RTE** is a digital publication, which can be accessed freely and openly from its website <http://www.rte.espol.edu.ec>. It is registered in several indexers and database directories and has a continuous evaluation of new forms of indexing that give greater visibility to published articles and their authors.

RTE presents scheduled calls for article submission through the Open Journal System platform, guaranteeing the publication of at least two volumes per year. However, each year additional volumes are produced and guaranteed on a specific subject (such as a Monograph or Dossier) or special editions related to a scientific or technological event, which collects articles or papers reviewed and accepted by blind peers before publication.

Editorial

*La Revista Tecnológica ESPOL - RTE, tiene el agrado de acoger como un volumen especial de sus ediciones 2023, la publicación de los artículos técnicos presentados en la **décima primera edición de la Conferencia Ecuatoriana de Tecnologías de la Información y Comunicación, TICEC**. Este evento de carácter científico y técnico es testimonio de la tenacidad y el dinamismo con que el Ecuador sigue trazando caminos innovadores en el vasto campo de la tecnología. Celebrada en la histórica y culturalmente rica ciudad de Cuenca del 18 al 20 de octubre de 2023, en las instalaciones de la Universidad de Cuenca, esta edición nos muestra una vez más la profundidad y amplitud de temas que están a la vanguardia de nuestra comunidad académica y profesional.*

Desde la amalgama entre sostenibilidad y tecnología evidenciada en el diseño de estaciones de carga solares, hasta la aplicación de la tecnología en el estudio y mejoramiento de nuestros recursos naturales como se ve en el análisis de la Cuenca del río Paute. Trabajos que reflejan la urgencia y la aplicación de conocimientos con que nuestros investigadores buscan respuestas técnicas a problemas complejos. Así mismo, los desafíos medioambientales evidenciados en estudios sobre contaminantes y el análisis climático, son retos que se abordan en conjunto con los avances en la industria tecnológica con esquemas disruptivos basados en inteligencia artificial, como la implementación y revisión del ChatGPT, demostrando una lucha por mantener el balance entre el respeto a nuestra tierra y la ambición de desarrollo tecnológico.

En esta edición, también se da un enfoque particular a temas sociales emergentes que tienen que ver con la tecnología que nos rodea en el día a día. Así, se analizan temas de ciberacoso, evidenciando la responsabilidad y conciencia social que los profesionales de la tecnología tienen sobre la vida cotidiana, inmersos en un mundo cada vez más digital. El compromiso con la educación y el bienestar se destaca en trabajos centrados en la estimulación de habilidades auditivas en niños, así como, en la atención y memoria cognitiva a través de juegos. Se subraya una vez más que la tecnología en manos adecuadas puede ser una herramienta extraordinaria para el bienestar de nuestra sociedad.

Este volumen especial de RTE, es más que una simple compilación de trabajos académicos, es un reflejo de nuestra identidad como comunidad, una que busca integrar, comprender y aportar a mejorar nuestro entorno natural con el desarrollo tecnológico. Los contenidos publicados en este volumen demuestran que nuestros profesionales se encuentran en la actualidad digital mundial, siempre a la vanguardia de la tecnología.

Agradecemos sinceramente a todos los autores, expositores y participantes quienes han hecho del TICEC 2023 un evento lleno de conocimiento, intercambio y visión de futuro. Esperamos que este volumen especial, Vol. 35, Núm. 2 de RTE, sirva como inspiración y fuente de referencia para todos aquellos interesados en el amplio mundo de las tecnologías de la información y comunicación.

Nayeth Solórzano Alcívar, Ph.D.
Editora en Jefe RTE

Santiago Berrezueta, Ph.D.
Editor Asociado RTE

Villie Morocho, Ph.D.
Chair del Track Técnico TICEC

Conferencia Ecuatoriana de Tecnologías de la Información y Comunicación, TICEC 2023
Cuenca, Ecuador
Octubre de 2023

Editorial

The Journal “*Revista Tecnológica ESPOL – RTE*”, is pleased to welcome a special volume of the 2023 edition, the publication of the technical articles presented in the **eleventh edition of the Ecuadorian Conference on Information and Communication Technologies, TICEC**. This scientific and technical event is testimony to the tenacity and dynamism with which Ecuador continues to chart innovative paths in the vast field of technology. Held in the historic and culturally rich city of Cuenca from October 18 to 20, 2023, in the facilities of the University of Cuenca, this edition shows us once again the depth and breadth of topics that are at the forefront of our academic and professional community.

From the amalgam between sustainability and technology evidenced in the solar charging stations to the application of technology in the study and improvement of our natural resources as seen in the Paute River Basin. These works reflect the urgency and the application of knowledge with which our researchers seek technological answers to complex problems. Likewise, the environmental challenges evidenced in the study of pollutants and climate analysis, are challenges that are addressed in conjunction with advances in the technology industry with disruptive schemes based on artificial intelligence, such as the implementation of ChatGPT, demonstrating a struggle to maintain the balance between respect for our land and the ambition of technological development.

In this edition, a particular focus is also given to emerging social issues that have to do with technology that surrounds us daily. Thus, cyberbullying issues are analyzed, evidencing the responsibility and social awareness that technology professionals have in everyday life, immersed in an increasingly digital world. The commitment to education and well-being is highlighted in our world focused on the stimulation of auditory skills in children, as well as attention and cognitive games. It is underlined once again that technology in the right hands can be an extraordinary tool for the well-being of our society.

This special volume of *RTE* is more than just a compilation of academic papers, it is a reflection of our identity as a community, one that seeks to integrate, understand, and contribute to improving our natural environment with technological development. The contents published in this volume demonstrate that our professionals are at the forefront of technology, always at the forefront of the digital world.

We sincerely thank all the authors, exhibitors, and participants who have made **TICEC 2023** an event full of knowledge, exchange, and vision for the future. We hope that this special volume, Vol. 35, Num. 2 of *RTE*, will serve as an inspiration and source of reference for those of all interested in the wide world of information and communication technologies.

Nayeth Solórzano Alcívar, Ph.D.
Editor in Chief RTE

Santiago Berrezueta, Ph.D.
Associate Editor RTE

Villie Morocho, Ph.D.
Chair of the TICEC Technical Track

Ecuadorian Conference on Information and Communication Technologies, TICEC 2023
Cuenca, Ecuador
October 2023

Edición Especial: TICeC2023

- 12 *Diseño y construcción de un prototipo de estación de carga solar para dispositivos móviles / Design and construction of a prototype of a solar charging station for mobile devices*
- 27 *Comparación de metodologías para la predicción del caudal mediante los índices de teleconexión. Caso de estudio: Cuenca del río Paute / Comparison of methodologies for flow prediction through teleconnection indices. Case study: Paute river basin*
- 50 *Análisis de la Calidad de las Asociaciones de Contaminantes y Variables Meteorológicas por Estacionalidad / Analysis of the Quality of the Associations of Pollutants and Meteorological Variables by Seasonality*
- 61 *Metodología para localizar la Zona de Convergencia Intertropical usando velocidad de viento / Methodology to locate the Intertropical Convergence Zone using wind velocity*
- 76 *El modelo OSeMOSYS y la descarbonización a largo plazo del sector eléctrico en Galápagos / OSeMOSYS model and the long-term decarbonization of the power sector in Galapagos*
- 94 *Diseño de un Software para Apoyar tareas de Mantenimiento de Equipos de Laboratorio Clínico / Design of a Software to Support Clinical Laboratory Equipment Maintenance*
- 109 *Aplicación web para la estimulación de las habilidades auditivas en niños / Web app for the stimulation of auditory skills in children*
- 122 *Modelo de Visualización de Datos de Juegos Serios de Atención y Memoria Cognitiva / Data Visualization Model of Serious Games of Cognitive Attention and Memory*
- 143 *Modelo Visual del Comercio Externo en Exportaciones Ecuatorianas / Visual Model of Foreign Trade in Ecuadorian Export*
- 157 *Análisis de Clusterización en Datos de Encuestas sobre Ciberacoso / Clustering Analysis on Cyberbullying Survey Data*
- 170 *Algoritmo para Clasificar la Resolución de Conflictos en Aspirantes Automotrices / Algorithm to Classify Conflict Resolution among Automotive Applicants*
- 181 *Generación de Texto Guía para la Detección Automatizada del Acoso y el Ciberacoso / Text Generation Guide for Automated Detection of Bullying and Cyberbullying*
- 192 *Síntesis Textual de Evaluación para Acoso y Ciberacoso / Textual Synthesis of Test for Bullying and Cyberbullying*



**LAS TIC
Y SU
IMPACTO
EN LA
SOCIEDAD**

**TRACK
TÉCNICO**

RTE

Diseño y construcción de un prototipo de estación de carga solar para dispositivos móviles

Design and construction of a prototype of a solar charging station for mobile devices

Miguel Alejandro¹ <https://orcid.org/0009-0000-4369-6800>, Rodolfo Merino² <https://orcid.org/0000-0003-2713-0588>, Jorge Carrión³ <https://orcid.org/0000-0003-1869-4541>, Andy Vega⁴ <https://orcid.org/0000-0003-0106-6880>, Jonathan González² <https://orcid.org/0000-0002-2207-8322>, Alba Vargas⁴ <https://orcid.org/0000-0003-0633-2316>

¹Universidad Nacional de Loja, Loja, Ecuador
miguel.alejandror@unl.edu.ec

²Centro de Investigación I²TEC, Loja, Ecuador
rodolfo.merino@unl.edu.ec, jmgonzaleze@unl.edu.ec

³Carrera de Electricidad, Universidad Nacional de Loja, Loja, Ecuador
jorge.carrion.g@unl.edu.ec,

⁴Carrera de Telecomunicaciones, Universidad Nacional de Loja, Loja, Ecuador
andy.vega@unl.edu.ec, alba.vargas@unl.edu.ec



Esta obra está bajo una licencia internacional
Creative Commons Atribución-NoComercial 4.0

Enviado: 2023/08/05

Aceptado: 2023/08/28

Publicado: 2023/10/15

Resumen

En esta investigación se realizó el diseño y construcción de un modelo de estación de carga solar para dispositivos móviles con el apoyo del Centro de Investigación “I²TEC”, de la Universidad Nacional de Loja (UNL). El trabajo inició con la revisión bibliográfica de los elementos de un sistema de energía solar fotovoltaico para definir conceptos y procedimientos que aporten al diseño del prototipo. Luego, se llevó a cabo el dimensionamiento teórico de los componentes necesarios y se seleccionaron los dispositivos cuyas especificaciones técnicas se asemejaban más a las disponibles en el mercado. El diseño final del sistema se implementó en una estructura de acero galvanizado y acero inoxidable, por la resistencia a la corrosión y fuerzas físicas

Sumario: Introducción, Metodología, Resultados y Discusión y Conclusiones.

Como citar: Alejandro, M., Merino, R., Carrión, J., Vega, A., González, J. & Vargas, A. (2023). Diseño y construcción de un prototipo de estación de carga solar para dispositivos móviles. *Revista Tecnológica - Espol*, 35(2), 12-26. Recuperado a partir de <http://www.rte.espol.edu.ec/index.php/tecnologica/articulo/view/1015>

que presentan estos materiales. Posteriormente, se realizó la instalación del prototipo propuesto en los exteriores de la Facultad de la Energía, las Industrias y los Recursos Naturales No Renovables (FEIRNNR) de la UNL para que sea utilizado por los estudiantes. Finalmente, se desarrollaron pruebas para verificar el correcto funcionamiento del prototipo y manuales de uso y mantenimiento del sistema para su adecuada conservación.

Palabras clave: Cargador, fotovoltaico, energía renovable, panel solar, radiación.

Abstract

In this research, a model of a solar charging station for mobile devices was designed and constructed with the support of the “I²TEC” Research Center of the National University of Loja (UNL). The work began with a bibliographic review of the elements of a photovoltaic solar energy system to define concepts and procedures that contribute to the prototype design. Then, the theoretical dimensioning of the necessary components was carried out, and the devices with the technical specifications closest to those available on the market were selected. The final system design was implemented in a galvanized steel and stainless steel structure due to the resistance to corrosion and the physical strength of these materials. The proposed prototype was then installed outdoors at the Faculty of Energy, Industries, and Non-Renewable Natural Resources (FEIRNNR) at UNL for student use. Finally, tests were developed to verify the correct operation of the prototype, and user manuals and maintenance guides were developed to ensure proper system preservation.

Keywords: Charger, photovoltaic, renewable energy, solar panel, radiation.

Introducción

Entre las formas de energía generadas por fuentes limpias, la solar fotovoltaica destaca en todo el mundo ya que no emite gases de efecto invernadero u otros contaminantes que produzcan daños al entorno. A diferencia de los recursos de petróleo, carbón y gas natural, esta energía no es finita. Si se utiliza cuidadosamente y en las aplicaciones adecuadas, puede proporcionar un suministro energético fiable y sostenible de forma casi indefinida (Alexis & Cedenno, 2020; Coria, 2014; Dincer & Rosen, 2013; Mani & Pillai, 2010; Toshov & Saitov, 2019).

Diversos acuerdos nacionales e internacionales establecen que este tipo de energía será uno de los principales contribuyentes para lograr un suministro 100 % renovable (Sierra & Reinders, 2021). Además, gracias al desarrollo tecnológico, su costo se ha reducido considerablemente, haciéndola más asequible para hogares y empresas. Al colocar paneles solares en cubiertas de instalaciones, ya sean viviendas, sistemas de riego, redes de telecomunicaciones, etc., las personas pueden producir electricidad y reducir su dependencia de las compañías eléctricas y de los combustibles fósiles (Aparicio, 2020; Dolezal et al., 2013; Lan et al., 2020; Setyawati, 2020). Esta energía también puede suministrar electricidad a regiones remotas que no tienen acceso a sistemas eléctricos tradicionales (Borges Neto et al., 2010; Lal et al., 2011; Zoulias & Lymberopoulos, 2007).

Estudios en este ámbito evidencian que el rendimiento de los módulos solares ha mejorado significativamente y se espera que siga mejorando en el futuro. También, se ha investigado la integración de acumuladores energéticos con paneles solares (Nasiri, 2008). Con ello se espera que aumente la eficiencia energética y se reduzcan costos. Además, existen nuevas formas de fabricar paneles solares con materiales más baratos y sostenibles, como las células

solares orgánicas y las perovskitas (Arabpour Roghabadi et al., 2018; Battersby, 2019; Sharma et al., 2015). En general, la investigación sobre energía fotovoltaica pretende incrementar su desempeño y reducir los costos para que esta tecnología sea más accesible y sostenible en todo el mundo; así como caracterizar adecuadamente a este sistema energético para evaluar y controlar su rendimiento (Jordehi, 2016).

Emplear tecnología solar fotovoltaica en puntos de carga de dispositivos móviles se ha convertido en una tendencia que crece rápidamente. Estas estaciones son una solución atractiva para brindar acceso eléctrico a poblaciones de bajos ingresos económicos, que usan poca energía y se encuentran en zonas remotas sin conexión eléctrica (Almeida & Brito, 2015) o instituciones en donde se requiere brindar comodidad para cargar tales equipos. Suelen estar conformadas por celdas solares, baterías y unidades que permiten cargar dispositivos, como: celulares, tablets, portátiles, etc. (Rosales Perez, 2019). Dichas celdas convierten la luz proveniente del sol en electricidad que luego se almacena en el dispositivo acumulador para su posterior uso. Cuando se necesita cargar un dispositivo, se enchufa el cargador y se utiliza la electricidad que ha sido almacenada.

Las personas destinan aproximadamente una o dos horas diarias para cargar sus dispositivos móviles y, en ocasiones, no disponen de suministro eléctrico o terminales de carga, especialmente en ambientes al aire libre (Grisales et al., 2018). Mantenerse informados constantemente y conectados con otras personas, ha hecho de los dispositivos móviles una parte fundamental de la vida cotidiana.

En la FEIRNNR - UNL se ha identificado la falta de espacios físicos que permitan a estudiantes y docentes desarrollar sus actividades académicas en un ambiente al aire libre con acceso a energía para cargar dispositivos móviles, ya que todos los tomacorrientes se encuentran ubicados en aulas y laboratorios. Asimismo, se ha constatado que estos puntos de carga existentes son tradicionales porque utilizan energía eléctrica convencional. Dado que en esta facultad se estudian e investigan las fuentes de energía limpia, y en colaboración con el Centro de Investigación, Innovación, Desarrollo de Prototipos Tecnológicos y Pruebas Electroenergéticas "I²TEC", se ha planteado el diseño y construcción de un modelo de estación de carga solar. El propósito de este prototipo es brindar a estudiantes, docentes y a quienes acceden a la facultad, la posibilidad de disponer de un espacio equipado para la recarga de dispositivos móviles, y que esté alineado con los principios de energías sustentables.

El procedimiento empleado para determinar los parámetros de diseño del sistema energético, así como los criterios de selección de los componentes y el presupuesto necesario para fabricar el prototipo, se describen en la sección de Metodología. En la sección de Resultados y Discusión, se detallan los logros obtenidos tras la instalación del sistema propuesto. En el apartado de Discusión, se presenta el análisis de dichos resultados y se verifica si se ha alcanzado el objetivo de la investigación. Finalmente, se exponen las conclusiones y recomendaciones derivadas de la ejecución de la propuesta.

Metodología

En esta sección se describe el procedimiento empleado para el desarrollo del proyecto. Se han definido tres componentes fundamentales: el diseño del sistema de energía, la implementación del prototipo y la elaboración de manuales destinados a guiar el uso y mantenimiento apropiado del sistema.

Diseño del sistema de energía

Para diseñar el sistema de energía del prototipo propuesto se realizó la proyección del consumo esperado, así como los cálculos necesarios para definir cada uno de los componentes del mismo.

Proyección de consumo

El sistema se diseñó para cargar dispositivos móviles con un consumo mínimo de 4 W y un consumo máximo de 20 W. Se consideraron seis tomas para el terminal de carga solar: cuatro de corriente continua tipo USB-B y dos tomas de corriente alterna de tipo B estándar para dispositivos que no cuentan con adaptador USB.

El valor de potencia necesaria se obtuvo multiplicando la potencia media consumida (0,012 kWh) por el número total de dispositivos y las horas que estarán conectados al sistema. Asumiendo una disponibilidad de 08:00 a 21:00 (horario laboral de la UNL), es decir, 14 horas y, que al menos la mitad de las tomas se encuentren conectadas durante este tiempo, o sea 3 dispositivos móviles, la potencia es de 0,504 kWh.

Elección de panel solar

Se evaluó la radiación solar en el área de implementación. Los datos utilizados se derivaron de dos fuentes de información. En primer lugar, se recurrió al Mapa Solar del Ecuador correspondiente al año 2019, en el cual se obtuvo el valor de la radiación promedio mensual producida a lo largo de ese periodo, así como los valores mínimos y máximos (Vaca Revelo & Ordóñez, 2020). La segunda fuente de información provino de la base de datos incorporada en el software *Solarisus PV*, utilizado para el modelado de instalaciones fotovoltaicas. Mediante esta fuente, se registró la localización específica de la zona de implementación del sistema solar y se extrajo el valor de radiación solar promedio mensual para el sitio en cuestión. El análisis de los datos recolectados de ambas fuentes permitió calcular un promedio y reveló que la irradiación más baja se presenta en el mes de junio (4,71 kWh/m²), mientras que la más alta se registra en octubre (5,48 kWh/m²).

Para determinar el número de paneles solares que deben utilizarse, se aplicó la ecuación (1), donde N_T es el número total de paneles, E_T es la energía usada diaria (Wh), P_{MPP} es la potencia nominal del módulo fotovoltaico y HSP es la Hora Pico Solar. Cabe destacar que esta ecuación se deriva de la metodología usada en el estudio “Atlas Solar del Ecuador con fines de Generación Eléctrica” (Corporación para la Investigación Energética CIE, n.d.).

$$N_T = \frac{E_T}{P_{MPP} (HSP)} \quad (1)$$

La HSP se refiere a la cantidad de energía solar disponible en un área. Una HSP es una hora en la que la intensidad solar es de 1 000 W/m² (Solar Power Authority Staff, n.d.). Si se toma el valor de irradiación promedio más baja (mes de junio 4 710 Wh/m²), se tiene la ecuación (2) (SESLab - Tecnológico de Costa Rica, n.d.):

$$HSP = \frac{\text{Irradiación}_{\text{DiariaPromedio}}}{1000 \text{ W/m}^2} \quad (2)$$

$$HSP = \frac{4710 \text{ Wh/m}^2}{1000 \text{ W/m}^2} = 4,71 \text{ h} \quad (3)$$

Tomando el valor de HSP conforme a la ecuación (3), el consumo diario de energía de 504 Wh y la potencia nominal de 110 Wp del módulo RSM110P, la aplicación de la ecuación (1) arroja el resultado de un módulo solar requerido para el sistema.

$$N_T = \frac{504 \text{ Wh}}{110 \text{ Wp}(4,71 \text{ h})} = 0,972 \approx 1 \text{ módulo solar} \quad (4)$$

Las celdas del módulo RSM110P son del tipo policristalino y están configuradas por 36 celdas (4x9). Posee una eficiencia del 16 % y cumple con las certificaciones ISO 9001, TUV PID-FREE, CE IEC61215/ 61730/61701/62716 (Resun Solar Energy, n.d.).

Controlador de carga

Para su selección se consideró el valor de corriente máxima de cortocircuito del generador fotovoltaico. Ya que el módulo solar elegido, modelo RSM110P, posee una corriente de 6,26 A en circuito cerrado, el controlador seleccionado debe poseer una tolerancia mayor a esa corriente. Se eligió el regulador modelo SHS-10 del proveedor *MorningStar* de 10 A, 12 VDC y bajo consumo de 8 mA (Morningstar Corporation, n.d.).

Inversor de corriente DC/AC

Es importante tomar en cuenta la potencia total para la elección del inversor. Para el cálculo, se consideró el valor máximo (20 W) de los dispositivos móviles que podrían conectarse de manera simultánea (6), obteniendo un total de 120 W. La potencia del dispositivo de conversión debe ser mayor al valor calculado, por ello se eligió el modelo TW-500 de 500 W de la marca ZONHAN (ZONHAN, n.d.).

Batería

Para la selección de la batería apropiada para el sistema se realizó el cálculo de su capacidad total, expresado por la ecuación (5) (AREATECNOLOGIA, n.d.). Se evaluó la demanda máxima (D_{max}), los días de autonomía (D_a), la profundidad de descarga (P_d) y el voltaje del sistema ($V_{sistema}$).

$$C_{Bat} = \frac{D_{max}(D_a)}{P_d(V_{sistema})} \quad (5)$$

Se determinó que la demanda máxima se alcanzará al conectar simultáneamente los 6 dispositivos, equivaliendo a 120 Wh. Se estimó un tiempo promedio de independencia de 4 horas, en las cuales no existirá luz solar. La profundidad de descarga es del 50 %, según lo indicado por el fabricante. El voltaje del sistema es de 12 V, conforme lo establece el regulador de carga. En base a estos valores, se determinó la necesidad de una batería de 80 Ah. Sin embargo, para asegurar un margen superior, se eligió la batería de gel modelo HTB12-100, la cual ofrece una capacidad de 100 Ah y un voltaje de 12 VDC (ProViento SA., 2023).

Tomacorrientes

El terminal de carga fue diseñado para seis puntos de conexión: cuatro de ellos corresponden a tomas de corriente directa tipo USB-B, mientras que los dos restantes corresponden a tomas de corriente alterna tipo B, que se subdividen en dos unidades. Para cumplir con este criterio, se eligió el modelo de toma polarizada que incorpora 2 tomas USB y 1 toma de corriente alterna tipo B, del fabricante Maviju. Este modelo presenta una tensión de 125 – 250 V, corriente 15 A + 2 USB (2,1 A – 1 A).

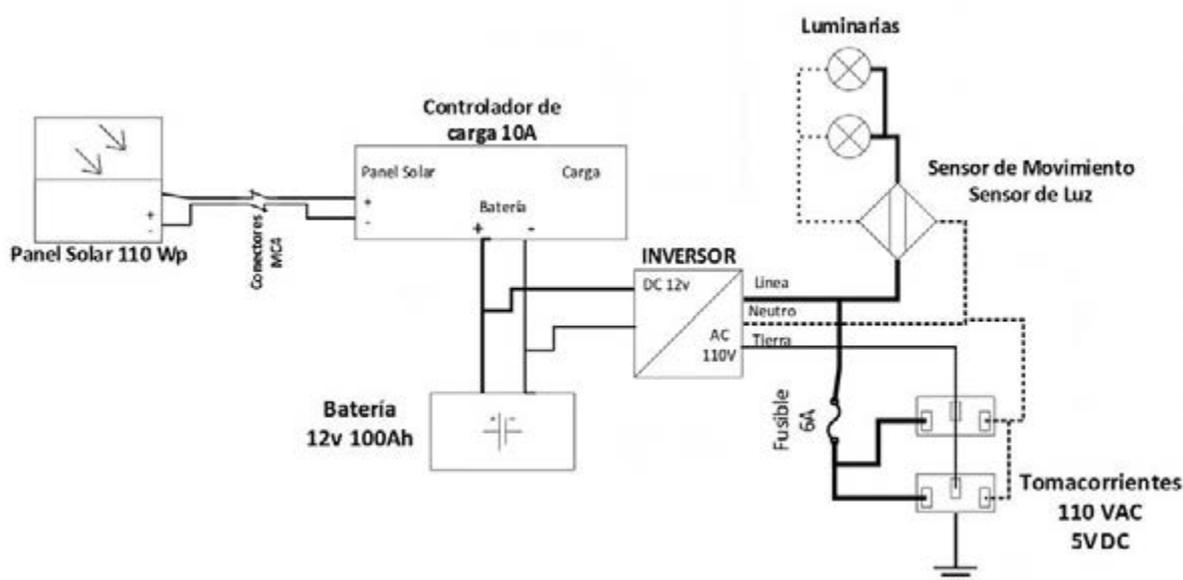
Sistema de iluminación

Para complementar la capacidad de carga de dispositivos móviles, se construyó un sistema de iluminación inteligente con luces LED, controlados por un sensor crepuscular y de movimiento para que encienda automáticamente las luces al existir movimiento y ante la ausencia de luz. Las luminarias seleccionadas son del tipo Panel cuadrado de 6 W, con un rango de voltaje de 100 – 240 V y una luminosidad de 350 lm de la marca Sylvania. El sensor de movimiento es de 360° de cobertura y opera en un rango de voltaje de 120 – 240 V, suministrado por el fabricante Maviju.

Diagrama eléctrico

En la Figura 1 se observa el diagrama de conexiones eléctricas diseñado para el modelo de carga.

Figura 1
Diagrama Eléctrico



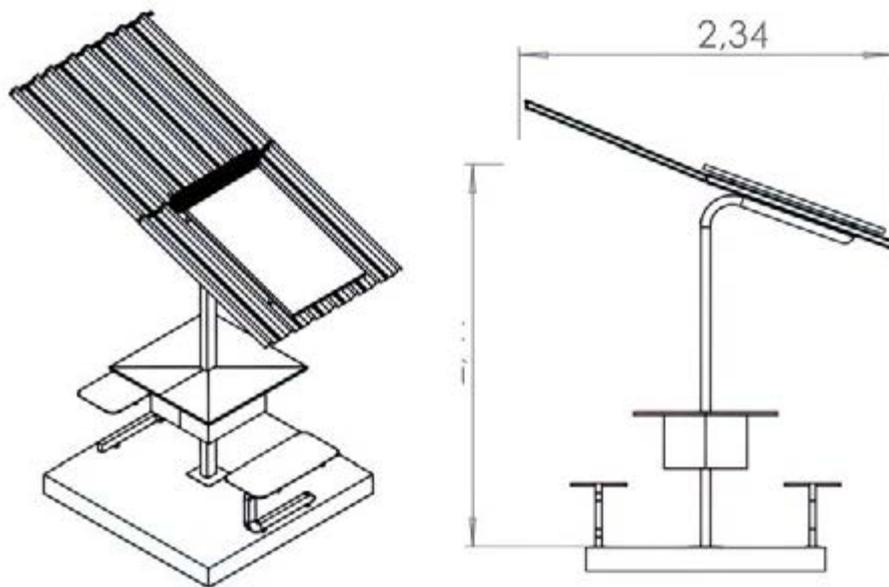
Para la conexión del panel se usó cable de calibre 8 AWG, mientras que para las demás conexiones se utilizaron cables de calibre 10 y 14 AWG. Como medida de protección, se empleó un fusible de 6 A, acorde a la potencia máxima de consumo establecida en el diseño. Además, se incorporó una puesta a tierra con una resistencia de 5 Ohm, con el propósito de salvaguardar los dispositivos móviles que se conectarán al sistema.

SopORTE estructural

Dado que el prototipo será instalado en los exteriores de la Facultad de la Energía - UNL, la selección de los materiales adquiere gran importancia. Estos deben ofrecer resistencia ante factores como la corrosión y las tensiones físicas provocadas por la intemperie. Deben ser capaces de proteger los componentes del prototipo, a la vez que proporcionen un lugar cómodo y funcional para estudiantes, docentes y el público en general.

En la Figura 2 se observan distintas vistas del diseño realizado, utilizando software especializado para modelado mecánico.

Figura 2
Diseño de Modelo de Carga



Con base en lo previamente expuesto, la estructura se construyó usando acero galvanizado y acero inoxidable. Consta de un eje principal formado por dos tubos de acero galvanizado soldados entre sí. En su parte superior, posee una superficie que sirve como cubierta y soporte para el panel solar. En el tramo intermedio están alojados los tomacorrientes que serán usados por los usuarios. Se colocó un compartimiento que alberga el resto de componentes del sistema. Además, se diseñaron dos bancos situados frente a cada enchufe, con el fin de proporcionar comodidad al usuario (Figura 3).

Figura 3
Propuesta en 3D de Soporte Estructural para Estación de Carga



Implementación

El prototipo fue instalado en la Facultad de la Energía - UNL, cuyas coordenadas geográficas son:

- Latitud: 4° 1'50.23" S
- Longitud: 79°11'58.88" O

Desarrollo de manuales

Los manuales de uso y mantenimiento parten desde una breve descripción de la estación solar y el objetivo del diseño. Se realizó un diagrama de conexiones con el propósito de brindar información técnica detallada sobre los elementos que conforman el sistema. Asimismo, se abordaron los problemas que eventualmente pueden ocurrir, junto con la solución propuesta para afrontarlos. Finalmente, se estableció una secuencia de pasos específicos para la realización del mantenimiento del sistema, asegurando así un correcto funcionamiento.

Presupuesto

En la Tabla 1 se desglosa el costo total empleado para la construcción del modelo propuesto. Todos los materiales y dispositivos utilizados fueron adquiridos en comercios locales y nacionales.

Tabla 1
Presupuesto para Construcción del Modelo

ÍTEM	MATERIAL - DISPOSITIVO	CANTIDAD	UNIDAD	VALOR UNITARIO (\$)	VALOR TOTAL (\$)
1	Panel solar	1	Unidad	80,00	80,00
2	Controlador de carga	1	Unidad	67,20	67,20
3	Batería	1	Unidad	291,20	291,20
4	Inversor	1	Unidad	112,00	112,00
5	Sensor de movimiento	1	Unidad	7,00	7,00
6	Luminaria LED	2	Unidad	5,00	10,00
7	Tomacorriente polarizado	2	Unidad	20,00	40,00
8	Cable 10 AWG	10	Metro	1,50	15,00
9	Cable 14 AWG	10	Metro	1,00	10,00
10	Cable 8 AWG	5	Metro	2,00	10,00
11	Material menudo	1	Global	40,00	40,00
12	Estructura metálica	1	Global	400,00	400,00
13	Mano de obra	1	Global	400,00	400,00
				TOTAL	1.482,40

Resultados y Discusión

En este apartado se presentan los hallazgos derivados de la investigación, así como su análisis y la evaluación del sistema propuesto.

El sistema energético solar diseñado e implementado se caracteriza por su autonomía, ya que no se halla conectado, ni depende de otra fuente de energía que no sea la producida por sí mismo. Su composición abarca un módulo fotovoltaico, un acumulador, un controlador de carga y un convertor de corriente.

El generador solar transforma la irradiación obtenida del sol en electricidad. El regulador o controlador de carga limita la tensión de salida para evitar la sobrecarga de la batería; además, regula su funcionamiento cambiando su modo operativo entre carga y reposo, dependiendo de la producción energética del módulo fotovoltaico, y activándose en función de las condiciones. La batería sirve como reserva energética, permitiendo la operación del sistema cuando el módulo solar no genera energía suficiente por sí mismo. Por su parte, el inversor convierte la corriente continua producida por el dispositivo fotovoltaico y la batería en corriente alterna, lo que habilita su uso en equipos convencionales.

En la Tabla 2 se describen los resultados del dimensionamiento del sistema.

Tabla 2
Dimensionamiento del Sistema Fotovoltaico

PARÁMETRO DE DISEÑO	CANTIDAD
Potencia demandada	0,504 kWh
Radiación promedio más baja	4710 Wh/m ²
Hora Solar Pico	4,71 h
Número de módulos solares de 110 WP	0,972
Potencia máxima necesaria	120 W
Potencia máxima generada	500 W
Capacidad de batería requerida	80 Ah
Capacidad de batería real	100 Ah

La Tabla 3 proporciona un desglose de los equipos utilizados en la construcción de la estación de carga solar.

Tabla 3
Equipos de Sistema de Carga Solar

ELEMENTO DEL SISTEMA	MODELO	CARACTERÍSTICAS GENERALES
Panel solar	RSM110P	110 Wp, Policristalino, 36 celdas
Controlador de carga	SHS-10	10 A y 12 VDC
Batería	HTB12-100	100 Ah, 12 VDC
Inversor	TW-500 ZONHAN	500W, IN:12V, OUT:110V
Sensor de movimiento	360 ° 50 MM Maviju	120-240 V, ajustable de 3 lx a 2.000 lx
Luminaria LED	Panel cuadrado 6 W Sylvania	100-240 V, 350 lm, 6 W
Tomacorriente	Toma polarizado Maviju	125-250V, 15A+2 USB (2.1A –1A)

En la Figura 4 se observa el prototipo implementado siendo utilizado por los estudiantes de la facultad durante la noche.

Figura 4
Estación de Carga Implementada



Para comprobar que el modelo de carga solar funcionara correctamente, se conectaron distintos dispositivos móviles y se verificó que se cargaran correctamente. En la Figura 5 se muestra la toma de medidas para validar los resultados, los cuales se detallan en la Tabla 4.

Tabla 4
Voltajes de Salida de Dispositivos

EQUIPO	VOLTAJE
Panel solar	21,50 V DC
Controlador de carga	12,78 V DC
Batería	12,78 V DC
Inversor	115,50 V AC
Tomacorriente AC	114,90 V AC
Tomacorriente USB	5,13 V DC

Figura 5
Toma de Medidas de Voltaje del Sistema



En base a la bibliografía disponible, se calcularon teóricamente los valores energéticos que el sistema debía entregar a los usuarios; se consideró un escenario en el que se alcance el número máximo de usuarios que usan el terminal de carga en un periodo de tiempo específico.

Para la selección de los materiales y dispositivos requeridos para construir el prototipo, se hicieron cálculos teóricos para identificar los equipos cuyas características técnicas se aproximaban más a las disponibles comercialmente.

Los materiales para armar el soporte estructural del prototipo se seleccionaron según su capacidad para resistir tanto la corrosión como las fuerzas físicas causadas por el clima. Tras un análisis, se determinó que los materiales más idóneos para ofrecer una mayor protección eran el acero galvanizado y el acero inoxidable.

Previamente a la instalación del sistema, se analizó que la zona elegida cuente con una superficie adecuada que permita realizar trabajos adicionales de mejora del suelo para garantizar una correcta fijación de la estructura.

Una vez instalado el prototipo, se probó el sistema midiendo el voltaje a la salida de cada tomacorriente, comprobando el correcto funcionamiento de cada equipo y del sistema en su conjunto.

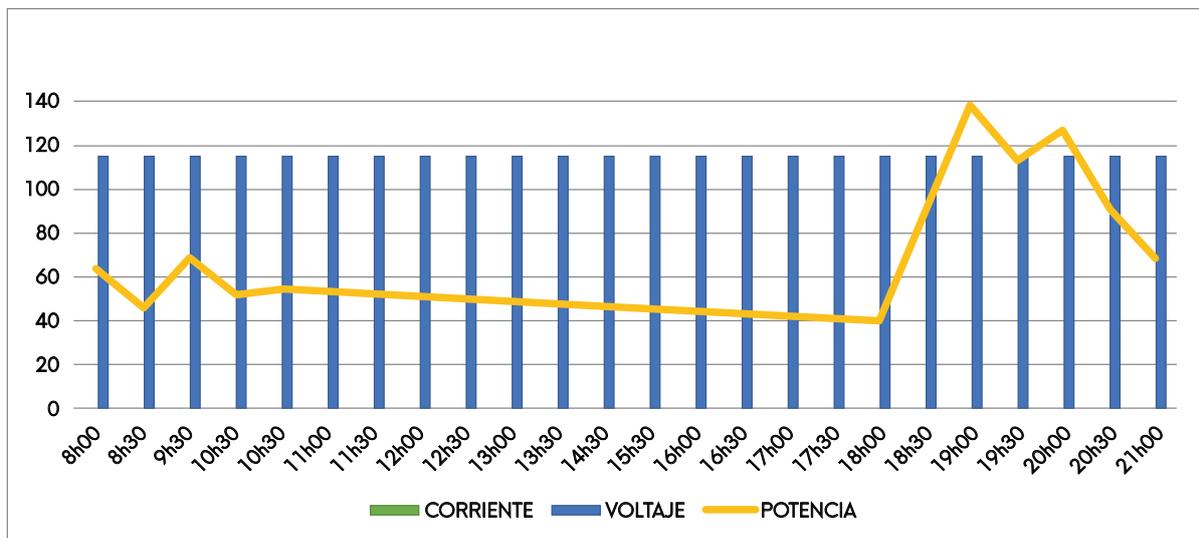
Como prueba de verificación final, se solicitó a los estudiantes que conecten sus dispositivos móviles al prototipo, y se logró observar que se cargaron correctamente (Figura 6).

Figura 6
Conexión de Dispositivos Móviles al Prototipo



La Figura 7 ilustra el monitoreo que se efectuó mediante el equipo FLUKE 435-2, en el cual se observó el uso en términos de corriente y potencia requeridos por los usuarios del prototipo. En esta representación se evidencia la concordancia de los datos calculados y los datos reales promedio de consumo.

Figura 7
Consumo Energético Promedio del Prototipo Implementado



Distintos criterios se tomaron en cuenta para elaborar el manual de uso y mantenimiento de la estación; por ejemplo: conceptos generales de los dispositivos que componen el sistema, su objetivo de diseño, posibles problemas que puedan ocurrir durante su funcionamiento, soluciones a dichos inconvenientes y recomendaciones para conservar apropiadamente el prototipo.

En lo que respecta al impacto tecnológico y social, la creación de espacios que permitan visualizar las aplicaciones y ventajas de las energías verdes dentro del campus universitario, pretende concientizar tanto a los estudiantes como a los docentes sobre la importancia de este tipo de fuentes energéticas para la sostenibilidad de las actividades humanas. Adicionalmente, se establece que este modelo podría ser replicado por gobiernos municipales para crear espacios verdes y tecnológicos que la ciudadanía pueda usar libremente.

Conclusiones

Se diseñó, construyó e implementó un modelo de estación de carga que emplea energía solar con el propósito de ofrecer a estudiantes, docentes y usuarios en general de la Facultad de Energía – UNL un medio para recargar sus dispositivos móviles y a la vez brindar un espacio cómodo para el desarrollo de sus actividades académicas al aire libre.

El análisis comparativo de información proveniente de diferentes fuentes bibliográficas sobre los elementos que componen un sistema fotovoltaico permitió la formulación y aplicación de conceptos y recomendaciones necesarias para la correcta implementación del punto de carga solar.

La selección de los equipos y accesorios del sistema se llevó a cabo mediante el uso de cálculos teóricos basados en fórmulas derivadas de la revisión bibliográfica. Estos valores obtenidos permitieron dimensionar correctamente cada elemento del prototipo y obtener el desempeño esperado.

Inspeccionar previamente la zona de implementación del modelo propuesto fue fundamental para la selección de materiales capaces de resistir el impacto de la intemperie y de las fuerzas físicas que puedan presentarse. Esta evaluación también permitió garantizar la protección de los componentes internos y brindar un lugar cómodo para los estudiantes. Como resultado de este análisis, se concluyó que utilizar acero galvanizado y acero inoxidable era una alternativa favorable.

La revisión de diversas fuentes bibliográficas sobre la irradiación solar en el lugar escogido para instalar el modelo, reveló que el valor promedio más bajo se registra en el mes de junio, con 4,71 kW/m², mientras que el valor más alto se observa en el mes de octubre, alcanzando 5,48 kW/m².

Realizar mediciones a la salida de los dispositivos instalados; así como, durante su funcionamiento, permitió determinar que el sistema en conjunto funciona correctamente y según las expectativas establecidas.

Estudiantes, docentes y demás personal de la Facultad de Energía - UNL pueden utilizar el prototipo de carga solar para acceder a electricidad y recargar sus dispositivos móviles. Esto les permite experimentar las ventajas derivadas del uso adecuado de fuentes de energía verde y verificar sus aplicaciones prácticas.

La viabilidad del prototipo, diseñado para su implementación y replicación en zonas como campus universitarios, parques y zonas comunes, se sustenta en el presupuesto establecido y en la facilidad para la adquisición de los materiales y equipos utilizados.

En términos de trabajos futuros, se considera la posibilidad de diseñar una estación solar de carga para bicicletas eléctricas y monopatinés, así como la implementación de sistemas de iluminación solar mimetizada y sostenible. Estas iniciativas permitirán la creación de un ambiente que fomente la aplicación de energías limpias y promueva la concientización sobre la sostenibilidad en zonas urbanas.

Reconocimientos

Esta propuesta forma parte del Trabajo de Titulación: “Diseño y construcción de un prototipo de estación solar de carga de dispositivos móviles para el I²TEC de la Universidad Nacional de Loja”. Los autores expresan su agradecimiento a Miguel Alejandro González por su participación en el trabajo y al Centro de Investigación I²TEC de la UNL.

Referencias

- Alexis, B. L. R., & Cedeno, E. A. L. (2020). The generation of electrical energy for industrial development in Ecuador from the use of renewable energy. *Universidad Ciencia y Tecnología*, 24 (104 SE-ELECTRIC INGEENIERING). <https://doi.org/10.47460/uct.v24i104.364>
- Almeida, R. H., & Brito, M. C. (2015). A review of technical options for solar charging stations in Asia and Africa. *Aims Energy*, 3(3), 428–449.
- Aparicio, M. P. (2020). *Energía solar fotovoltaica: 3a edición*. Marcombo.
- Arabpour Roghabadi, F., Ahmadi, N., Ahmadi, V., Di Carlo, A., Oniy Aghmiuni, K., Shokrolahzadeh Tehrani, A., Ghoreishi, F. S., Payandeh, M., & Mansour Rezaei Fumani, N. (2018). Bulk heterojunction polymer solar cell and perovskite solar cell: Concepts, materials, current status, and opto-electronic properties. *Solar Energy*, 173, 407–424. <https://doi.org/https://doi.org/10.1016/j.solener.2018.07.058>
- AREATECNOLOGIA. (n.d.). *Dimensionado Instalación Fotovoltaica*. <https://www.areatecnologia.com/electricidad/dimensionado-instalaciones-fotovoltaicas.html>
- Battersby, S. (2019). The solar cell of the future. *Proceedings of the National Academy of Sciences*, 116(1), 7–10. <https://doi.org/10.1073/pnas.1820406116>
- Borges Neto, M. R., Carvalho, P. C. M., Carioca, J. O. B., & Canafístula, F. J. F. (2010). Biogas/photovoltaic hybrid power system for decentralized energy supply of rural areas. *Energy Policy*, 38(8), 4497–4506. <https://doi.org/https://doi.org/10.1016/j.enpol.2010.04.004>
- Coria, I. D. (2014). Energías renovables en la provincia de Santa Fe: situación actual y perspectivas. *Invenio: Revista de Investigación Académica*, 33, 5–7.
- Corporación para la Investigación Energética CIE. (n.d.). *Atlas solar del Ecuador con fines de generación eléctrica*. https://www.academia.edu/40441973/ATLAS_SOLAR_DEL_ECUADOR_CON_FINES_DE_GENERACION_ELÉCTRICA_Corporación_para_la_Investigación_Energética
- Dincer, I., & Rosen, M. A. (2013). Chapter 4 - Exergy, Environment And Sustainable Development. In I. Dincer & M. A. Rosen (Eds.), *Exergy (Second Edition)* (Second Edi, pp. 51–73). Elsevier. <https://doi.org/https://doi.org/10.1016/B978-0-08-097089-9.00004-8>
- Dolezal, A., Majano, A., Ochs, A., & Palencia, R. (2013). La ruta hacia el futuro para la energía renovable en Centroamérica. *Evaluación de La Situación Actual, Mejores Prácticas, Análisis de Brechas*. Washington, USA: Worldwatch Institute.

- Grisales, E. A. D., Ruiz, J. D. G., Grisales, P. M. O., Tobin, A. F. L., Lopez, S. C., & Piedrahta, A. F. I. (2018). Design and Construction of a Stand-Alone PV System for Charging Mobile Devices in Urban Landscapes in Medellin. *International Journal of Renewable Energy Research (IJRER)*, 8(1), 258–265.
- Jordehi, A. R. (2016). Parameter estimation of solar photovoltaic (PV) cells: A review. *Renewable and Sustainable Energy Reviews*, 61, 354–371. <https://doi.org/https://doi.org/10.1016/j.rser.2016.03.049>
- Lal, D. K., Dash, B. B., & Akella, A. K. (2011). Optimization of PV/wind/micro-hydro/diesel hybrid power system in HOMER for the study area. *International Journal on Electrical Engineering and Informatics*, 3(3), 307.
- Lan, T. T., Jirakiattikul, S., Chowdhury, M. S., Ali, D., Niem, L. D., & Techato, K. (2020). The Effect of Retail Electricity Price Levels on the FI Values of Smart-Grid Rooftop Solar Power Systems: A Case Study in the Central Highlands of Vietnam. *Sustainability*, 12(21). <https://doi.org/10.3390/su12219209>
- Mani, M., & Pillai, R. (2010). Impact of dust on solar photovoltaic (PV) performance: Research status, challenges and recommendations. *Renewable and Sustainable Energy Reviews*, 14(9), 3124–3131. <https://doi.org/https://doi.org/10.1016/j.rser.2010.07.065>
- MORNINGSTAR Corporation. (n.d.). *SHS User's Manual*. <https://www.morningstarcorp.com/wp-content/uploads/operation-manual-shs-de-en-es-fr.pdf>
- Nasiri, A. (2008). Integrating energy storage with renewable energy systems. 2008 34th Annual Conference of *IEEE Industrial Electronics*, 17–18. <https://doi.org/10.1109/IECON.2008.4757918>
- ProViento SA. (2023). *Bateria Solar de GEL vida útil prolongada 100Ah/12VDC*. <https://proviento.com.ec/baterias-solares/199-bateria-solar-de-gel-vida-util-prolongada-100ah12vdc.html>
- Resun Solar Energy. (n.d.). *Small Modules—Off-Grid Modules*. Resun. <https://www.resunsolar.com/products/small-modules/>
- Rosales Perez, C. E. (2019). *Implementacion de un sistema de carga para celulares mediante energía solar fotovoltaica*.
- SESLab - Tecnológico de Costa Rica. (n.d.). 6. *El concepto de hora solar pico*. https://seslab.org/fotovoltaico/6_el_concepto_de_hora_solar_pico.html
- Setyawati, D. (2020). Analysis of perceptions towards the rooftop photovoltaic solar system policy in Indonesia. *Energy Policy*, 144, 111569. <https://doi.org/https://doi.org/10.1016/j.enpol.2020.111569>
- Sharma, S., Jain, K., & Sharma, A. (2015). Solar Cells: In Research and Applications—A Review. *Materials Sciences and Applications*, 1145–1155. <https://doi.org/10.4236/msa.2015.612113>
- Sierra, A., & Reinders, A. (2021). Designing innovative solutions for solar-powered electric mobility applications. *Progress in Photovoltaics: Research and Applications*, 29(7), 802–818. <https://doi.org/https://doi.org/10.1002/pip.3385>
- Solar Power Authority Staff. (n.d.). *How to Calculate Your Peak Sun-Hours*. Solar Power Authority. <https://www.solarpowerauthority.com/how-to-calculate-your-peak-sun-hours/>
- Toshov, J., & Saitov, E. (2019). Portable autonomous solar power plant for individual use. *E3S Web of Conferences*, 139, 1087.
- Vaca Revelo, D., & Ordóñez, F. (2020). *Mapa Solar del Ecuador 2019*.
- ZONHAN. (n.d.). *ZONHAN TW500 PURE SINEWAVE INVERTER*. https://www.proviento.com.pe/Zonhan_TW500_datasheet.pdf
- Zoulias, E. I., & Lymberopoulos, N. (2007). Techno-economic analysis of the integration of hydrogen energy technologies in renewable energy-based stand-alone power systems. *Renewable Energy*, 32(4), 680–696. <https://doi.org/https://doi.org/10.1016/j.renene.2006.02.005>

Comparison of methodologies for flow prediction through teleconnection indices. Case study: Paute river basin

Comparación de metodologías para la predicción del caudal mediante los índices de teleconexión. Caso de estudio: Cuenca del río Paute

María Daniela González¹ <https://orcid.org/0009-0003-3436-0805>, Darío Xavier Zhiña^{1,2} <https://orcid.org/0000-0001-9556-4025>, Alexandra Guanuchi-Quito^{1,2} <https://orcid.org/0000-0002-5583-8674>, Alex Avilés-Añazco^{1,2} <https://orcid.org/0000-0001-9278-5738>

¹Facultad de Ciencias Químicas, Universidad de Cuenca, Cuenca, Ecuador
daniela.gonzalez98@ucuenca.edu.ec

²Grupo de Evaluación de riesgos ambientales en sistemas de producción y servicios (RISKEN), Universidad de Cuenca, Cuenca, Ecuador
dario.zhina0608@ucuenca.edu.ec,
alexandra.guanuchi@ucuenca.edu.ec,
alex.aviles@ucuenca.edu.ec



Esta obra está bajo una licencia internacional
Creative Commons Atribución-NoComercial 4.0

Sent: 2023/06/24

Accepted: 2023/07/07

Published: 2023/10/15

Abstract

The Paute river basin (southern Ecuador) suffers hydrological changes due to climate change and human activities. Hydrological changes cause extreme events and affect ecosystems, hydroelectric plants, and quality of life. It highlights the importance of understanding hydrological behavior to make appropriate decisions in extreme environments. This study seeks to predict discharges in the Paute river basin through global teleconnection indices. Multiple Linear Regression (MLR) was obtained using three different methodologies: multicollinearity analysis, Principal Component Analysis (PCA), and correlation with monthly delays. It was shown that the principal component analysis scenario obtained the best predictive fits, specifically by including 41 indices and 20 components. For the scenario using monthly delays, the best delay occurs within a single month for most seasons. Finally, with the multicollinearity analysis scenario, better results were obtained using 41 indices, although essentially the performance corresponds to the number and

Summary: Introduction, Materials y Methods, Results and Discussion and Conclusions.

How to cite: González, M., Zhiña, D., Guanuchi-Quito, A., & Avilés-Añazco, A. (2023). Comparison of methodologies for flow prediction through teleconnection indices. Case study: Paute river basin. *Revista Tecnológica - Espol*, 35(2), 27-49.

Recuperado a partir de <http://www.rte.espol.edu.ec/index.php/tecnologica/article/view/1028>

indices of each model. Teleconnection indices are not sufficient when used as the only input variable for download modeling and prediction, giving mostly unsatisfactory results. However, a clear trend links the behavior of flows and indices, and it is possible to improve the models based on more climatic variables or with other predictive methods.

Keywords: Discharge prediction. Teleconnection indices. Principal component analysis. Multiple regression models. Multicollinearity analysis.

Resumen

La cuenca del río Paute (sur del Ecuador) sufre cambios hidrológicos por el cambio climático y las actividades humanas. Los cambios hidrológicos causan eventos extremos y afectan a ecosistemas, centrales hidroeléctricas y la calidad de vida. Destaca la importancia de comprender el comportamiento hidrológico para tomar decisiones adecuadas en ambientes extremos. Este estudio busca predecir las descargas en la cuenca del río Paute mediante los índices de teleconexión global. Se obtuvieron modelos de Regresión Lineal Múltiple (MLR) mediante tres metodologías diferentes: análisis de multicolinealidad, Análisis de Componentes Principales (ACP) y correlación con retrasos mensuales. Se demostró que el escenario de análisis de componentes principales obtuvo los mejores ajustes predictivos, específicamente al incluir 41 índices y 20 componentes. Para el escenario que usa retrasos mensuales, el mejor retraso ocurre dentro de un solo mes, para la mayoría de las estaciones. Finalmente, en el escenario de análisis de multicolinealidad se obtuvieron mejores resultados utilizando 41 índices, aunque esencialmente el rendimiento corresponde a la cantidad y los índices de cada modelo. Los índices de teleconexión no son suficientes cuando se utilizan como la única variable de entrada para el modelado y la predicción de descargas, dando resultados en su mayoría insatisfactorios. Sin embargo, existe una clara tendencia que vincula el comportamiento de caudales e índices, y es posible mejorar los modelos en base a más variables climáticas o con otros métodos predictivos.

Palabras clave: Bases de datos, predicción de caudales, Índices de Teleconexión, Análisis de Componentes Principales, Modelos de Regresión, Análisis de Multicolinealidad.

Introduction

Within studies related to global teleconnections between ENSO (El Niño South Oscillation) and discharge, strong and regionally consistent discharge impacts were found in Central and South America, New Zealand, and Australia, while weaker signals were observed in parts of Africa and North America (Kundzewicz, Szwed, & Pińskwar, 2019). To understand more the localized effects of teleconnections, a large number of studies are currently focused on connecting flooding with climatic variability on a continental scale. Many relevant studies have been carried out in Australia, Asia, Europe, North America, and South America. These studies utilize indices such as ENSO, NAO (North Atlantic Oscillation), AMO (Atlantic Multidecadal Oscillation), TSA (Tropical Southern Atlantic), TNA (Tropical Northern Atlantic), AO (Arctic Oscillation), and PDO (Pacific Decadal Oscillation), among others (Giddings and Soto, 2006).

In a local context, the most relevant index is ENSO due to its effects. In South America, ENSO causes floods and droughts along the western coast. However, simulating ENSO in the region faces biases and uncertainties, especially when dealing with long time series. Systematic errors occur in the central equatorial Pacific, the eastern equatorial Indian Ocean, and regions with boundary current systems (such as the tropical Pacific and Atlantic) (Cai et al., 2020).

Within the confines of the study area and its regional dynamics, some studies have

been conducted in Ecuador, demonstrating that the Sea Surface Temperature (SST) variability and ENSO phenomena have an impact on discharge patterns throughout the nation. Similarly, multiple investigations have established that discharge anomalies can be found nationwide using ENSO and its modes. (Córdoba Machado et al., 2015). The research holds particular significance due to its revelation that the three most common natural catastrophes in Ecuador are floods, droughts, and landslides, with the first two having a significant impact on local living conditions (Fontaine et al., 2008).

To comprehend the impacts of teleconnections on climate, it is imperative to situate these events within a climatological context. Teleconnection indices are statistically significant correlations of recurrent atmospheric anomalies that occur in nearby and distant areas, often concurrently, at the planetary or hemispheric level (Hatzaki et al., 2007). In line with this methodology, while examining the sequence of oscillations that constitute climate variability around the mean values, these anomalies are recognized by detecting deviations. This representation exemplifies the climate's ephemeral condition in response to modifications and is demarcated by precise temporal and spatial scales (IDEAM - UNAL, 2018). Additionally, the teleconnection fluctuations encompass anomalies within their cycles, generating variability modes. These modes are determined from sub-calculations or temporal variations in the behavior of teleconnection indices (Dima and Lohmann, 2004).

Broadly speaking, focusing on Ecuador and the Andes, ENSO is the primary factor influencing SST and air pressure. El Niño (warm phase) and La Niña (cold phase) patterns constitute this phenomenon. Another factor is the PDO, which exhibits warm and cold interdecadal phases that affect the surface waters of the Pacific Ocean. Recent investigations have identified that the impact of PDO in South America has increased due to climate change (Morán et al. 2016). Regarding precipitation, indicators related to ENSO (such as Niño 3.4 and SOI), can affect rainfall in Ecuador; also, Niño 3.4 and SST exhibit a significant causal relationship. The only significant impact that NP and WP have on rainfall in South America is in Brazil. Due to spatial distance, AO and AAO have a minimal influence. Widespread droughts have been related to TNA and TSA. However, droughts or excessive rains in Ecuador or the region show no significant connection to the NAO (Giddings and Soto, 2006).

Due to its ecological and social relevance, numerous hydrological studies have focused on regional factors within the Paute river basin in southern Ecuador to comprehend patterns in sub-basin and micro-basin discharge (Celleri et al., 2007; Sotomayor et al., 2018; Ward et al., 2011). However, considering the recent understanding of these teleconnection indices, the use of these to forecast water behavior and correlate it with extreme hydrological events represents a novel exploratory strategy. These indices operate on a global scale and might exert a greater influence on hydrological dynamics than local elements do.

As previously mentioned, the significance of this study stems from the fact that substantial ecological and human losses occur in the area, particularly due to floods in the Sierra region induced by rivers within the Paute River basin. For that reason, the purpose of the project is to use teleconnection indices and variability modes to forecast the flow in the Paute river basin. Principal component analysis (PCA), correlation with monthly delay, and MLR with multicollinearity analysis are all part of the process of forecasting. The levels of prediction for each methodology will then be evaluated using statistical measures by contrasting the outcomes with the original data.

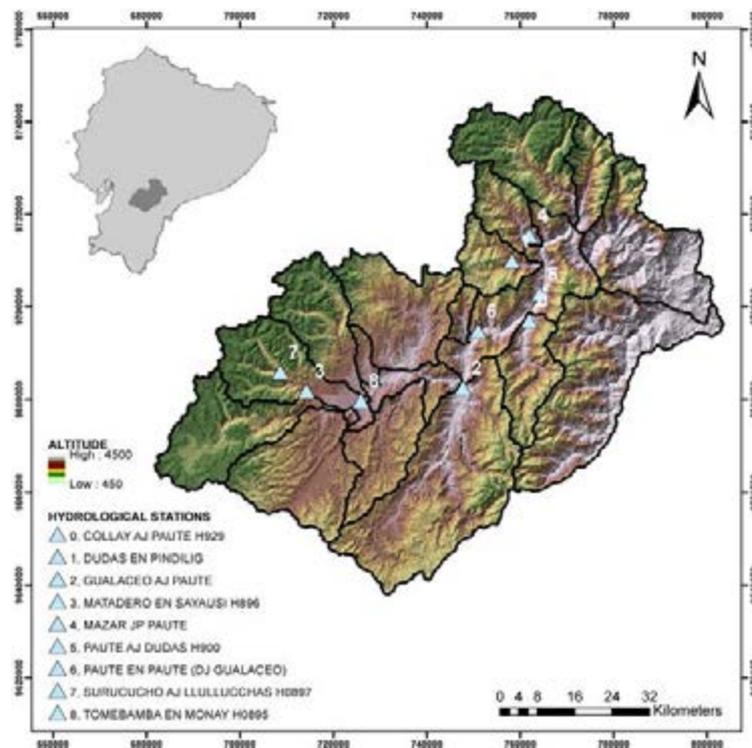
Materials y Methods

Study area

The study was conducted in the Paute river basin in southern Ecuador (Figure 1). The basin covers approximately 6437 km², with slopes ranging from 25% to 50% (CELEC EP 2013). The basin covers the Azuay, Cañar, and Morona Santiago provinces and is part of the Santiago River basin. Annual rainfall reaches its maximum average between 2500-3000 mm in the eastern region, while in the western region, it ranges between 1200-1500 mm (Institute of Regime Studies et al., 2017). The area experiences two distinct periods of frequent precipitation due to the ITCZ. The wet season for the unimodal regime occurs from June to August, while for the bimodal regime, it takes place from March to May (Campozano et al., 2016).

Figure 1

Location of the study site, as well as the location of the different sub-basins and the discharge and precipitation stations



The Paute basin consists of 18 hydrological sub-basins (Figure 1), which discharge, through a fall process, feeds the hydroelectric plants with a discharge value of Mazar (141.10 m³/s), Molino (200 m³/s), Sopladora (150 m³/s) and Cardenillo (180 m³/s), El Labrado and Chanlud (2.4 m³/s and 4.18 m³/s respectively) (Orbes & Peralta, 2017; Matute Pinos, 2014). The amount of energy produced by each of the aforementioned plants is Amaluza (1.075 MW), El Labrado and Chanlud dams feed two plants, Saymirin and Sucay (14.4 MW and 24 MW respectively); Mazar (162.6 MW) and Sopladora (500 MW). It is estimated that in total, it produces 40% of the hydroelectric production at the country level (Contreras et al., 2017).

Hydrometeorological and climatic data

Data from discharge stations of the Ecuadorian Institute for Meteorology and Hydrology (INAMHI) were collected for 20 years (1995-2015) on a monthly scale. Only stations with at least 40% data availability were considered, resulting in a total of nine stations (Table 1). Data filling was performed using the ten nearest discharge and precipitation stations for each selected station. The average monthly discharge amount was used to complete the missing data.

Table 1
Discharge and precipitation stations used for the study

STATION NAME	X	Y	ID	% MISSING VALUES
PAUTE AJ DUDAS	78.62	2.68	H900	53
GUALACEO AJ PAUTE	78.76	2.86	H0931	60
TOMEBAMBA EN MONAY	78.96	2.88	INA895	38
PAUTE EN PAUTE (DJ GUALACEO)	78.74	2.75	H894	35
SURUCUCHO AJ LLULLUCCHAS	79.12	2.83	INA897	18
MATADERO EN SAYAUSI	79.66	2.87	INA896	14
COLLAY AJ PAUTE	78.64	2.73	H929	50
DUDAS EN PINDILIG	78.68	2.62	H902_R	40
MAZAR JP PAUTE	78.64	2.57	H0892	59

In addition, data for time series of teleconnection indices and modes of variability were obtained from the database of the National Oceanic and Atmospheric Administration [41 indices (NOAA - <https://psl.noaa.gov/data/climateindices/list/>)].

Flow modeling using atmospheric and oceanic climatic indices

Teleconnection indicators, climate variability modes, and MLR were used in the flow modeling process. A variety of criteria, including PCA, correlation with monthly delays, and multicollinearity analysis, were used to generate MLRs. For calibration and validation, the study used 80% of the data. Data were utilized for calibration directly from January 1995 to September 2011, and for validation directly from September 2011 to December 2015.

Flow modeling using Principal Component Analysis (PCA)

PCA was used on teleconnection indices and climate variability modes to minimize the number of variables. MLR analyzed factors accounting for 70% and 90% of the variance (Rea and Rea 2016). R software's "prcomp" function was employed for this. PCA is essential for reducing dimensionality in data analysis while maintaining data variety and representativeness (Shabri and Shuhaida, 2014).

Modeling using Atmospheric and Oceanic Climate Indices and Monthly Delays

Three forecast scenarios were constructed to take into account the delay processes between atmospheric circulation and precipitation/flow effects. In these circumstances, the indices were delayed by the flow values by one, two, and three months. The top five teleconnection indicators for each discharge station with the highest Pearson correlation coefficients were chosen for MLR creation.

Modeling using multicollinearity analysis

A multicollinearity study was carried out to reduce the number of variables to address the problem of many variables. Multicollinearity, which indicates linear dependence across predictors and interferes with each predictor's effects on the dependent variable, was discovered using the Variance Inflation Factor (VIF) (Vega Vilca and Guzman, 2011). The appropriate VIF value was five, and variables that exceeded this limit were eliminated. For MLR building, the Leap Sequence criterion, combining forward and backward selections, was utilized (Hastie et

al., 2021). RStudio software with the MASS package and the "leapseq" function were utilized for this process.

Goodness of fit of the resulting models

To determine the accuracy of the predictions made, a series of metrics were processed in the R software V4.0.2, using the "hydroGOF" package. The metrics used in this study were KGE (Knoben, Freer and Woods, 2019), NSE (Krause, Boyle and Bäse, 2005), RMSE (Meyer, 2010), and R2. The model with the highest value in their metrics was considered the best model.

Results and Discussion

Modeling using Principal Component Analysis (PCA)

Of the 41 teleconnection indices, 10 and 20 main components were obtained, which explain 70% and 90% of the variance, respectively. It is considered that a 70% explanation of variance is a significant percentage for the study; however, the components that explained 90% of the explanation of variance were used to obtain a better prediction of flows (Rea and Rea, 2016).

The components obtained in the PCA analysis were used to perform a multiple regression model. The results obtained with the PCA analysis were plotted in time series and dispersion diagrams; the results can be seen in Figures 2 - 5.

Figures 2-5 illustrate that the results align with observed values, but predictions deviate from the actual values. Underestimation occurs at low values, while overestimation occurs at high values. Although the model performs better for lower values, it fails to accurately reflect extreme events. This limitation arises from using only MLR and station values without considering environmental behavior or hydrological processes.

Figure 2

Observed and modeled flow time series using MLR and ACP (calibration stage). a) Paute AJ Dudas, b) Gualaceo, c) Tomebamba, d) Paute DJ Gualaceo, e) Surucucho, f) Matadero, g) Collay, h) Dudas, i) Mazar

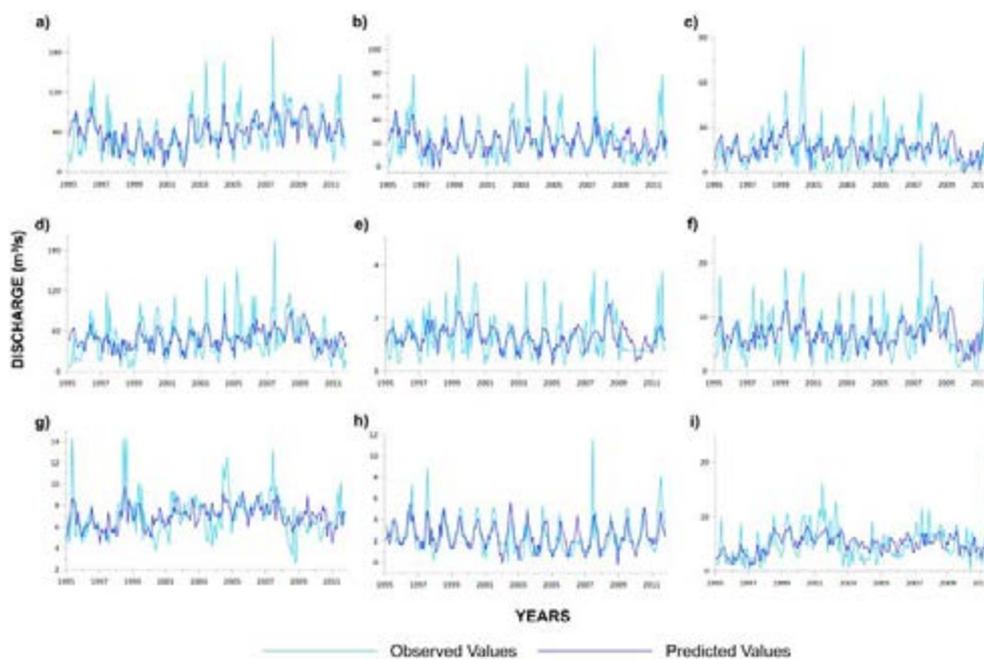


Figure 3

Observed and modeled flow time series using MLR and ACP (validation stage). a) Paute AJ Dudas, b) Gualaceo, c) Tomebamba, d) Paute DJ Gualaceo, e) Surucucho, f) Matadero, g) Collay, h) Dudas, i) Mazar

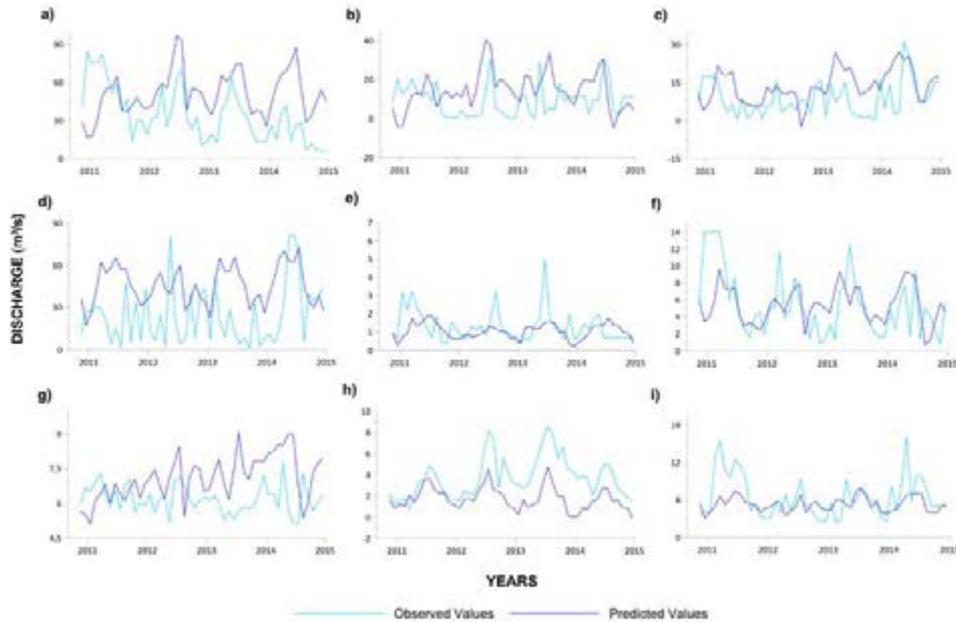


Figure 4

Dispersion plots of observed and modeled discharge using MLR and ACP (calibration stage).

a) Paute AJ Dudas, b) Gualaceo, c) Tomebamba, d) Paute DJ Gualaceo, e) Surucucho, f) Matadero, g) Collay, h) Dudas, i) Mazar

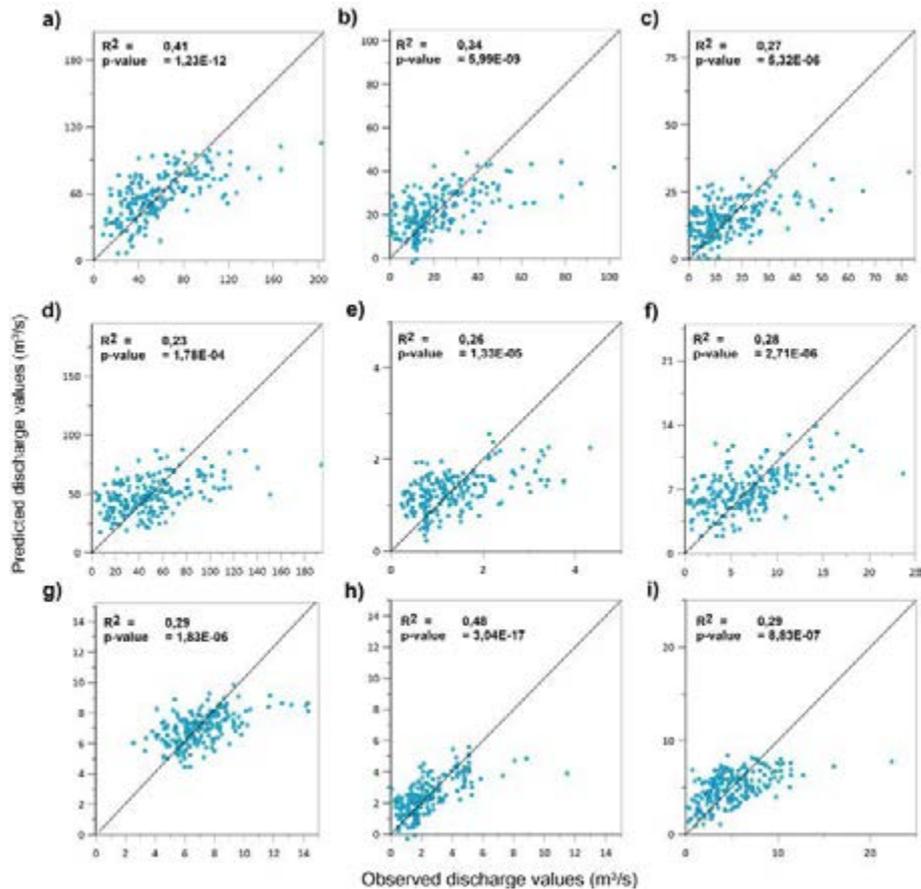
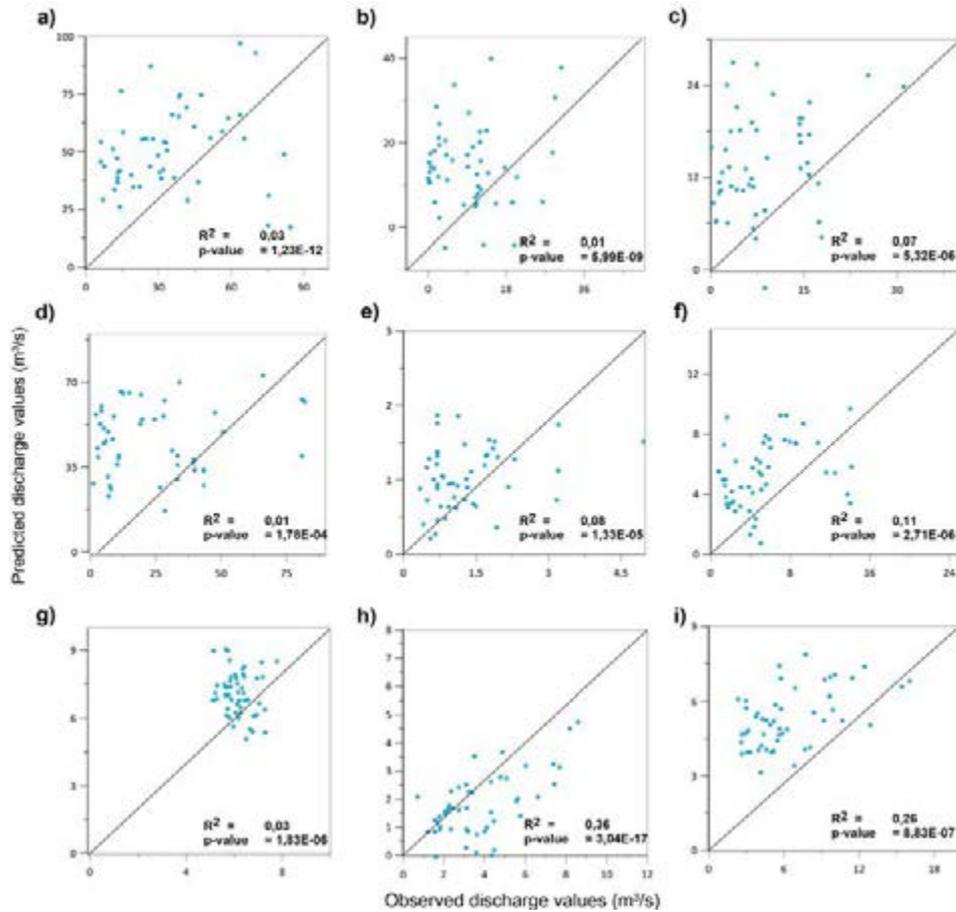


Figure 5

Dispersion plots of observed and modeled discharge using MLR and ACP (validation stage).

a) Paute AJ Dudas, b) Gualaceo, c) Tomebamba, d) Paute DJ Gualaceo, e) Surucucho, f) Matadero, g) Collay, h) Dudas, i) Mazar



During the validation phase, synchronization is seen in severe events that are within the station ranges of each participant. The precision is still less than at the calibration stage, though. The minimum p-values on the dispersion plots in both stages demonstrate statistical significance and a distinct correlation between the indices and the stations from the Paute basin. It's crucial to remember that these outcomes are dependent on the data utilized and independent of the robustness of the model.

Among all the stations, Paute AJ Dudas and Dudas (Figures 2a-2h, 3a-3h, 4a-4h, and 5a-5h) exhibit the best-fit models with correlations of 0.41 and 0.48, respectively, during the calibration stage. However, during the validation stage, none of the stations achieved satisfactory results, with the highest correlation observed in Dudas at 0.36 and Mazar at 0.26. This indicates that the models heavily rely on the initial conditions of the study and do not effectively incorporate other environmental or hydrological data.

Paute DJ Gualaceo and Surucucho demonstrate the lowest correlation results during the calibration stage, with values of 0.23 and 0.26 (Figures 2d-2e, 3d-3e, 4d-4e, and 5d-5e). Overall, the calibration results exhibit low correlations, sometimes as low as 0.01, indicating a lack of robustness in the models using this methodology.

Table 2
Metrics results at calibration and validation stage

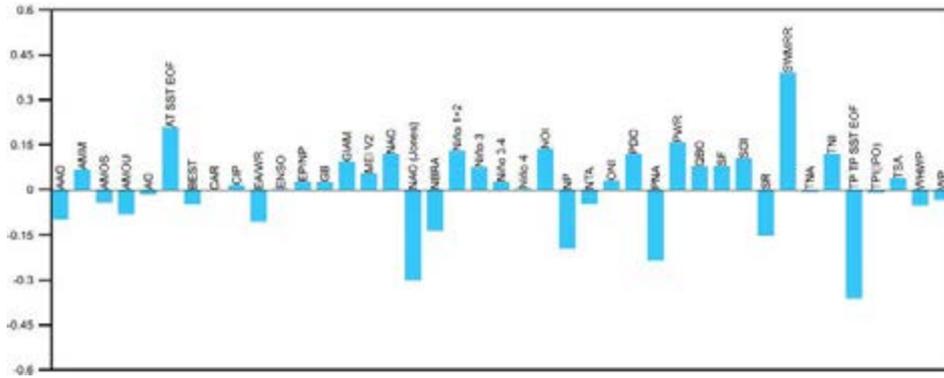
STATIONS	CALIBRATION STAGE				VALIDATION STAGE			
	RMSE	NSE	R2	KGE	RMSE	NSE	R2	KGE
PAUTE AJ DUDAS	24.35	0.41	0.41	0.49	30.25	-1.02	0.03	0.02
GUALACEO	13.27	0.34	0.34	0.41	12.93	-1.55	0.01	-0.06
TOMEBAMBA	10.74	0.27	0.27	0.33	9.5	-0.98	0.07	0.06
PAUTE DJ GUALACEO	26.05	0.23	0.23	0.27	32.07	-1.26	0.01	-0.29
SURUCUCHO	0.68	0.26	0.26	0.31	0.89	-0.05	0.08	0.08
MATADERO	3.56	0.28	0.28	0.34	3.67	0.05	0.11	0.2
COLLAY	1.6	0.29	0.29	0.34	1.52	-5.44	0.03	-0.33
DUDAS	1.19	0.48	0.48	0.57	2.5	-0.68	0.36	0.21
MAZAR	2.57	0.29	0.29	0.35	3.12	0.12	0.26	0.16

There aren't satisfactory results in these models; in Table 2 we can observe that Dudas, Collay, and Mazar were the stations with less RMSE even when only Dudas had a good model fit, which means this could be related to the low original discharge values and their predictions and not properly with the model functioning. According to Moriasi et al. (2007), the NSE does not show satisfactory results (from 0.6 or higher), classification describes a performance as very good, good, satisfactory, and unsatisfactory with values higher to 0.75, between 0.65 – 0.75, between 0.5 – 0.65 and lower to 0.5 respectively. Dudas and Paute AJ Dudas also domain this metric in the calibration stage. KGE gets better performance values than NSE, without changing the results of stations.

In the validation stage, Dudas still has one of the best performances, but Mazar and Matadero also appear here. These metrics do not have any special tendency. A study made in New Gales, Australia, using the indices ENSO, MEI, IOD, Niño 3.4, PDO, and TPI to predict the study zone discharges, got NSE values between 0.15 – 0.55, which means that indices by themselves are not enough to predict discharge (Esha, Imteaz y Nazari, 2019). A study realized in western Canada mentions that PDO (used in our study), even when related to ENSO, also has great uncertainty regarding the nature and origin of the index variability itself; which results in a waste of potential use in forecasting studies, especially because the statistic shows that it is highly related to water resources (Whitfield et al., 2010).

In comparison to other approaches, PCA had the lowest RSME, indicating that other indices may be more closely associated than the ENSO index or its variability modes (unique indices have the potential to affect only this methodology). Additionally, there may be several variables utilized for the study; however, the VIF approach challenges this idea because, in some instances, the indices used for the models soar to 13. Metrics show that incorporating 90% of variance in all models rather than just 70% results in an improvement.

Figure 6
Factorial Components of ACP using 41 indices and 20 Components



In Figure 6, we can observe the loading factor of ACP, which is the influence of every index on the 20 components. According to the study, the best models (90%) are indirectly related to ENSO indices and not to the study area, as can be identified in a comparison between the indices and their influence on each component used in the modeling (AT SST EOF, PWR, and SWMRR (South West Monsoon Region Rainfall) with positive values, NAO (Jones), NP, PNA, SR, and TP SST EOF (Tropical Pacific SST Empirical Orthogonal Function) (negatives). The highest correlations occur with SWMRR (0.39) and AT SST EOF (0.21). This correlation could occur because the rainfall season of SWMRR in Mexico and Arizona happens from June to September, matching the wet season of the Paute River Basin of the unimodal regimen from June to August (Campozano et al., 2016 & Crimmins, 2014). And AT SST EOF correlation is related to the location of this anomaly in Ecuador (Fan and Schneider, 2012).

The PNA and other ENSO patterns used in research on the Canadian Columbia River basin showed that depending on the station (certain months), ENSO variability modes could be more or less influent, even causing an anomaly in the river discharges (Gobena, Weber y Fleming, 2013). As the results of our models are not accurate, this can be considered an explanation, having indexes that are more related to anomalies in the ACP study, being necessary to realize a correlation study considering each station's anomaly values on the data time series.

In a study made in Iran, 25 variability modes of teleconnection indices were used to explain the variability of precipitation using ACP methodology (AMO, AMM, BEST, Niño 3.4, El Niño 4, NTA, SOI, and TNA). Results show eight principal components that explain 80% of the variance (Choubin et al., 2016). In Colombia, a study that analyzes the climatic variability of the Cauca River using the ACP in Index (CCC, an index of the own river; ONI; PDO; EMI, El Niño Midoki; SST; SOI; and MEI) has as results that two components that explain 80% of the variance, an only one explaining 70%. The eigenvalues (proportional to loading factors) showed that this first component with 70% of variance was more associated with ENSO and the two components with 80% of variance were more related to the index of the study zone CCC (Sedano, 2017). In our study, the models work similarly, most of the indices directly correlate with ENSO and SST. It is also visible in the studies of Choubin et al. (2016) and Sedano (2017) that the percentage of input data needed to explain 70% of the variance is from 25 – 30%, and for upper percentages, it is from 35 – 50% of them. The influence of every index involved decreases with the number of components, focusing on their correlation.

Modeling using Atmospheric and Oceanic Climate Indices and Monthly Delays

To determine the variables of each prediction model for the discharge stations, the Pearson correlation was carried out for the different stations and the teleconnection indices,

The MLRs were generated with the indices from Table 2. The results are presented in Figures 8-11 with a one-month delay; the behavior of the modeled time series adjusts to the trend of the series of each station, especially at mean values and low values. However, it presents the same problem as the series predicted with the PCA method. All values are indifferent to the model results.

Figure 8

Observed and modeled flow time series using MLR and one-month delay (calibration stage). a) Paute AJ Dudas, b) Gualaceo, c) Tomebamba, d) Paute DJ Gualaceo, e) Surucucho, f) Matadero, g) Collay, h) Dudas, i) Mazar

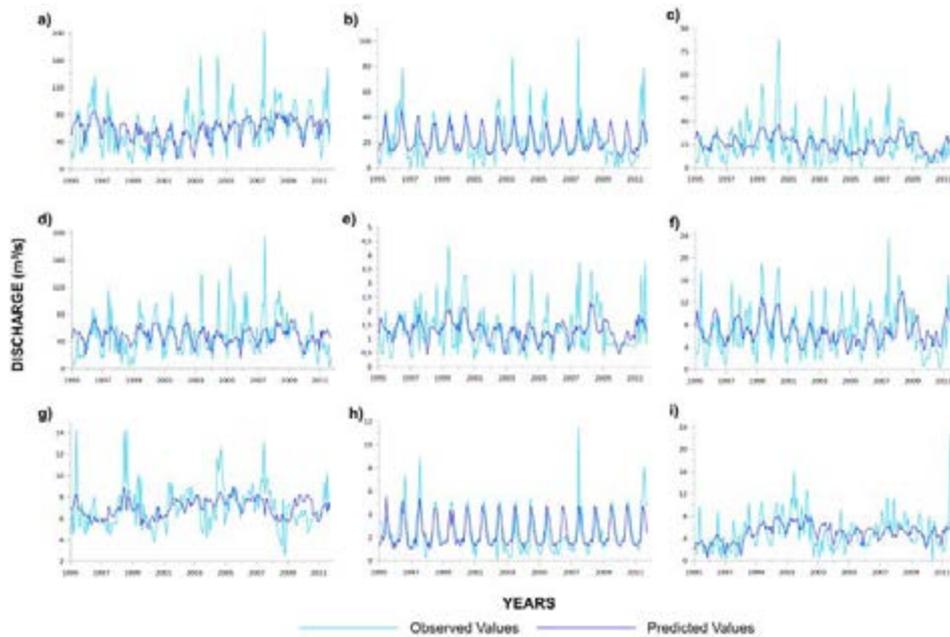
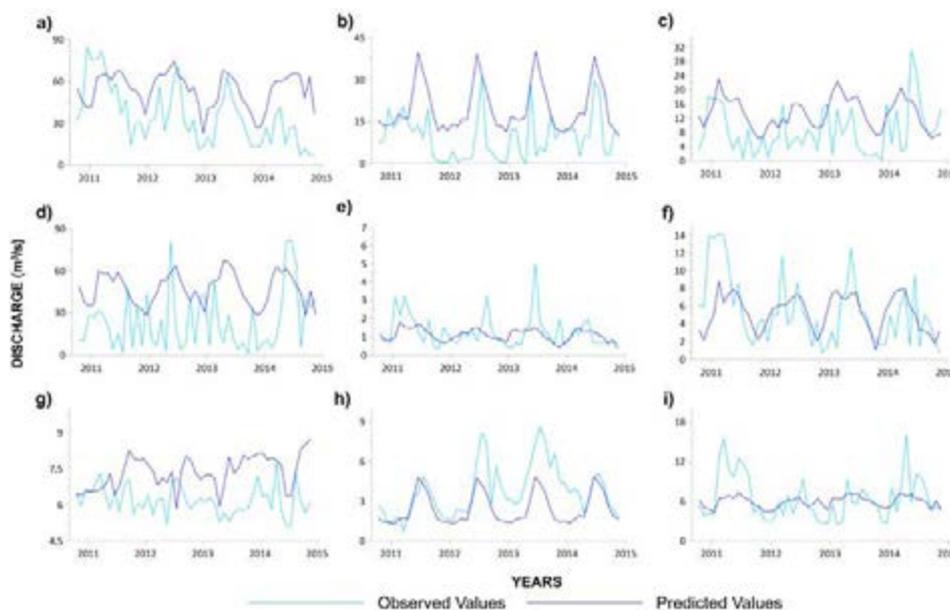


Figure 9

Observed and modeled flow time series using MLR and one-month delay (validation stage). a) Paute AJ Dudas, b) Gualaceo, c) Tomebamba, d) Paute DJ Gualaceo, e) Surucucho, f) Matadero, g) Collay, h) Dudas, i) Mazar



In 2018, an investigation into the evolution of the sediments of several lakes in the El Cajas National Park determined that, with a one-month delay for the ENSO index, the results of the correlation with the precipitations close to the local stations were solid, proving that there is a strong positive link between intense rains and therefore an increase in flow between La Niña, Niño 3.4 (which explains the variability of the Ecuadorian Andes), El Niño 4 (a more intense relationship), El Niño 1+2 (which explains the variability of the coastal plains), and a more neutral link with El Niño (Schneider et al., 2018).

The delay of one month in our study proves that there is a delay in the effects of the teleconnection rates on the climatic conditions of the Paute river basin. And that the variability of the indices fits with the level of correlation found considering the spatial configuration of the studies. In a study carried out in Brazil on the relationship between the alluvial plains and the connection with the signals (variability modes) of ENSO teleconnection (El Niño, Niño 3.4, SOI), it was determined that for the Amazon basin, there is a delay of two months. This occurs from January to March and causes lower rainfall, causing a decrease in flow (Schöngart et al., 2004). However, an investigation carried out in Peru on the relationship between monthly rainfall and SST identified that in the areas of the equatorial Pacific, ENSO has a delay of one month in the wet season and zero in the dry season (Bazo, Lorenzo and Porfirio Da Rocha, 2013). When comparing our results with the literature, it confirms our results because the indices with the highest correlation for this scenario are AMM, NP, PWR and, WHWP, which are strongly influenced by ENSO variability modes.

Figure 10

Dispersion plots of observed and modeled discharge using MLR and one-month delay (calibration stage). a) Paute AJ Dudas, b) Gualaceo, c) Tomebamba, d) Paute DJ Gualaceo, e) Surucucho, f) Matadero, g) Collay, h) Dudas, i) Mazar

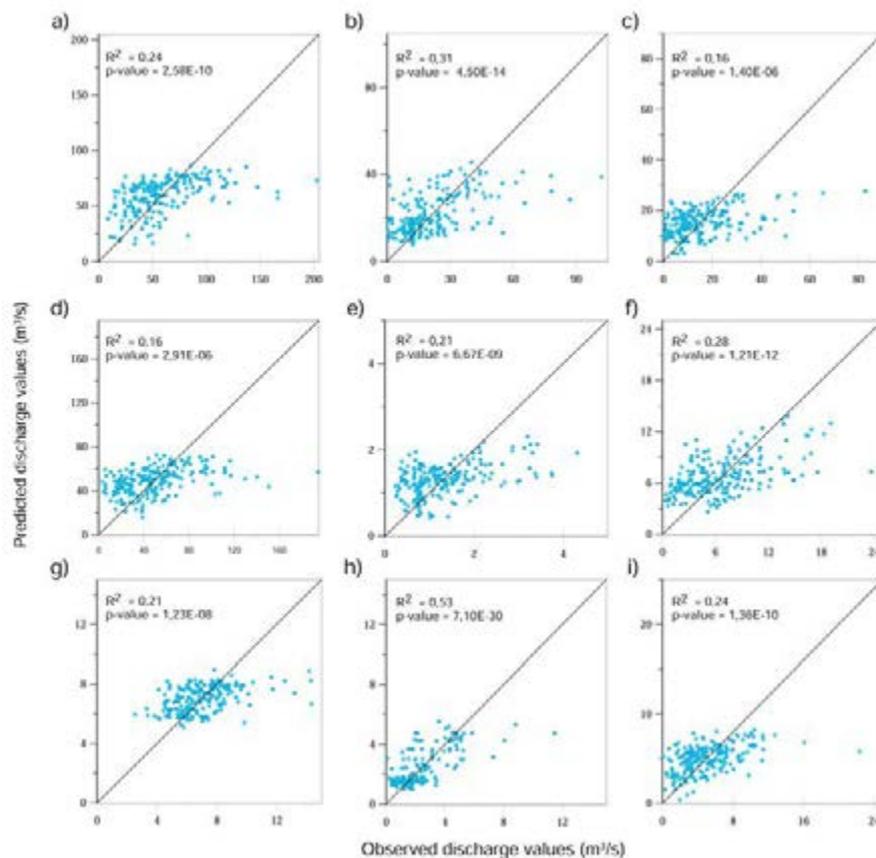
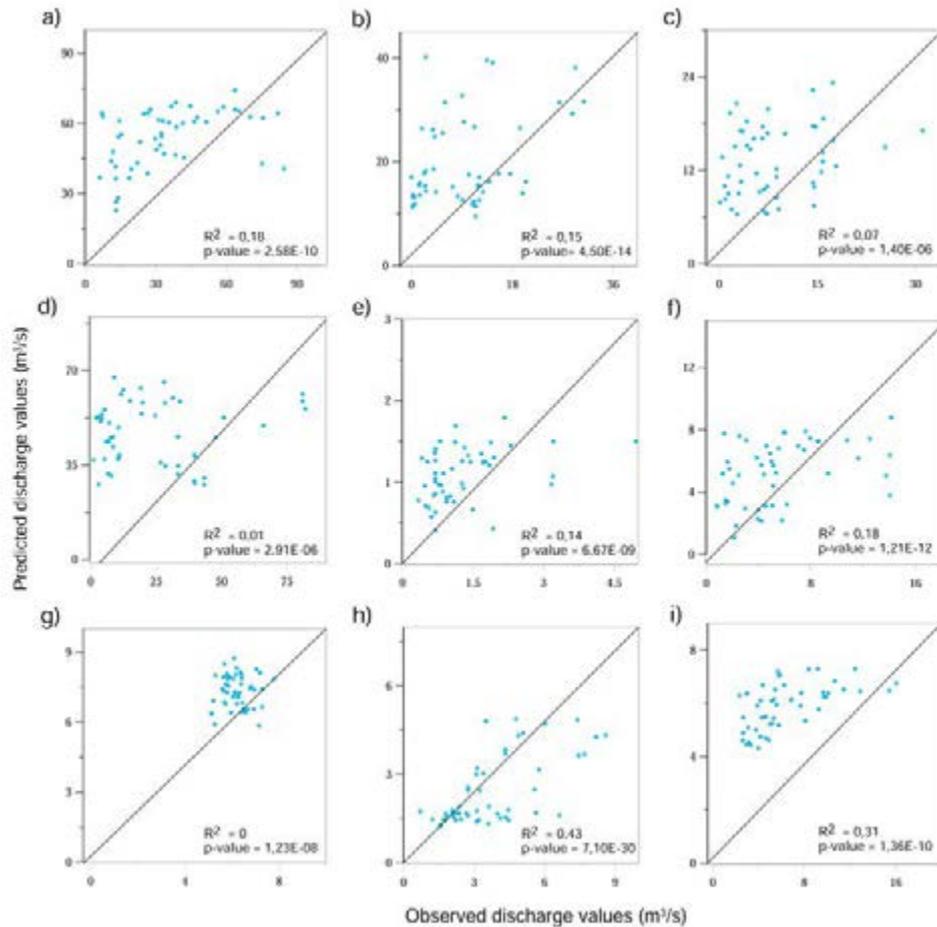


Figure 11

Dispersion plots of observed and modeled discharge using MLR and one-month delay (validation stage). a) Paute AJ Dudas, b) Gualaceo, c) Tomebamba, d) Paute DJ Gualaceo, e) Surucucho, f) Matadero, g) Collay, h) Dudas, i) Mazar



The PWR index is conditioned by El Niño (Carreric, 2020); on the other hand, the WHWP index is also directly affected by El Niño in the summer through the TNA temperature increase that occurs during the El Niño winter in the Pacific (Wang y Enfield, 2001). NP is dominated by the interannual variations of the El Niño and La Niña events (Espino Sánchez, 2014); AMM, for its part, produces anomalies in the SST north of Ecuador, but it occurs during the boreal spring (March) with greater intensity. In a case study of the rivers of Quebec in Canada was discovered that between ENSO, NAO, and PDO there is a large delay between the effect of the patterns and the answer on the discharge. It is also mentioned that the modes of variability associated with teleconnection patterns can change the periodicity and the stationary effects. This can explain the non-optimal fit models in this methodology (McGregor, 2017).

For the modeling, an acceptable R² value is obtained in Dudas with 0.53 (Figure 10h), which confirms the ability of the models to capture the averages. The second-best performance occurs at the Gualaceo station with a coefficient of 0.31 (Figure 10b) and Matadero with 0.28 (Figure 10f). In this case, two of the stations varied in the indices used, which represented a great improvement for both cases. The models with the lowest performance are Paute DJ Gualaceo and Tomebamba, both with a value of 0.16 (Figure 10c – 10d).

When analyzing the coefficient of determination (Table 4), it can be seen that Dudas station is acceptable according to the bibliography. Since values less than 0.5 indicate in a

hydrological model that there is a large error variance integrated into the model and that it cannot be explained by it, in other words, using only the values of the indices in a multiple regression model does not give good results for the flows of the Paute river basin. There are local and regional variables that could be integrated into the model, improving its predictive capacity (Moriassi et al., 2007).

Table 4
ACP metrics results at calibration and validation stage

STATIONS	CALIBRATION STAGE				VALIDATION STAGE			
	RMSE	NSE	R2	KGE	RMSE	NSE	R2	KGE
PAUTE AJ DUDAS	27.69	0.24	0.24	0.28	26.79	-0.65	0.18	0.11
GUALACEO	13.62	0.31	0.31	0.37	13.92	-1.95	0.15	-0.27
TOMEBAMBA	11.54	0.16	0.16	0.16	8.49	-0.6	0.07	0.01
PAUTE DJ GUALACEO	27.29	0.16	0.16	0.15	32.68	-1.37	0.01	-0.41
SURURCUCHO	0.71	0.21	0.21	0.24	0.84	0.08	0.14	0.09
MATADERO	3.57	0.28	0.28	0.34	3.44	0.16	0.18	0.25
COLLAY	1.69	0.21	0.21	0.23	1.46	-4.96	0	0
DUDAS	1.15	0.53	0.53	0.61	1.96	-0.06	0.43	0.37
MAZAR	2.66	0.24	0.24	0.28	2.98	0.2	0.31	0.13

In this modeling, we get the best fit of the model. NSE reaches a satisfactory result in the calibration stage with Dudas, and KGE gets a good fit model (0.61). Paute AJ Dudas's results are reduced compared with other methodologies; this can occur when analyzing where stations are located and how the delay effects affect every station. Tomebamba and Gualaceo have the lowest results, reaching 0.15 and 0.16 on NSE and KGE, respectively. RMSE is only related to the discharge of each station and not the results of the modeling.

In the calibration stage, Dudas is still the best station with 0.37 on KGE, while Collay gets the worst results with -4.96 on NSE and zero on KGE. This means that the average monthly values are a better predictor than our model (Krause et al., 2005). The variance range is very wide, and these results prove that the models have no statistical robustness even when the p-value maintains infinity in all cases (both stages).

Modeling using multicollinearity analysis

The third case used for flow modeling was a preliminary analysis of multicollinearity, as explained in the previous section on materials and methods. When using the VIF criterion for the reduction of the variables, a total of 22 variables were obtained (reduction of almost half of the original variables); the resulting variables were PNA, WP, EA/WR, NAO, TSA, PDO, NP, AO, AAO, PWR, CAR, AMOS, QBO, SR, SF, GB, EP/NP, NAO (Jones), NOI, CIP, NBRA, SWMRR and AT SST EOF. None of the modes of variability or signals directly related to ENSO (MEI V2, Niño 1+2, Niño 3, Niño 3.4, Niño 4, SOI, ONI) are reflected in the results. Indirectly, we find NP, which indicates its high influence on the rest of the indices and its intimately connected and similar behavior. This also happens with the indices AMO UNSMOOTHED, AMM, and AMO SMOOTHED, since only one is kept after VIF analysis. With the resulting

variables, the next step was the construction of the MLR using the stepwise criterion. The best model for each of the discharge stations is presented in Table 5.

Table 5
MLR used for prediction with Multicollinearity Analysis

DISCHARGE	N° VARIABLES	INDICES
PAUTE AJ DUDAS	5	AMOS + AT SST EOF + NP + PDO + SOLAR
GUALACEO	2	NP + PDO
TOMBAMBA	5	EA/WR + NP + PWR + CIP + NBRA
PAUTE DJ GUALACEO	4	AMOS + CAR + NP + PDO
SURUCUCHO	2	NP + PWR
MATADERO	1	PWR
COLLAY	2	CAR + SR
DUDAS	1	CIP
MAZAR	5	AMOS + NP + SOLAR+TSA+NBRA

Figures 12-15 show the results of the predictions made under the VIF analysis for multicollinearity. The trends are maintained in all seasons, and visually, the results are acceptable (Dudas achieves the best modeling).

Figure 12

Observed and modeled flow time series using MLR and Multicollinearity Analysis (calibration stage). a) Paute AJ Dudas, b) Gualaceo, c) Tomebamba, d) Paute DJ Gualaceo, e) Surucucho, f) Matadero, g) Collay, h) Dudas, i) Mazar

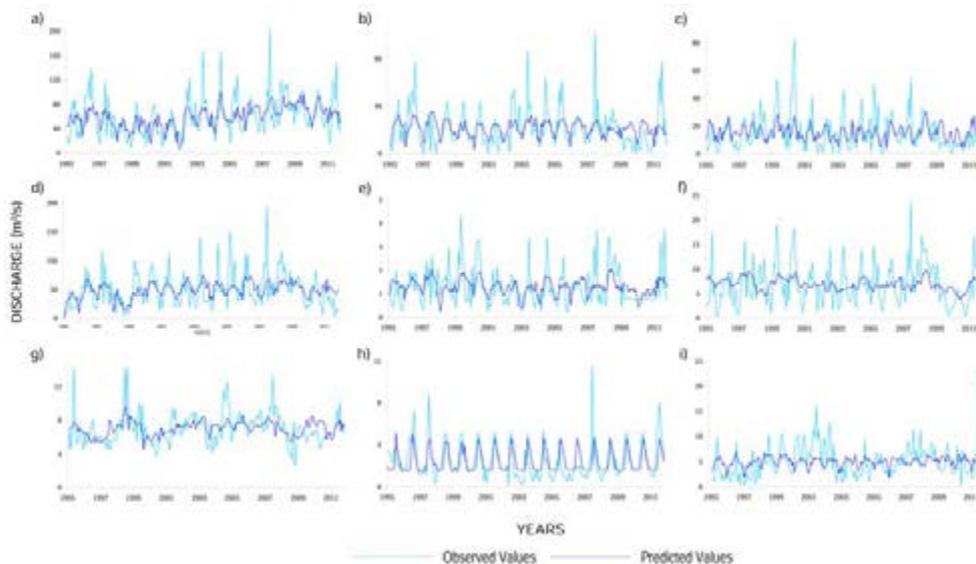
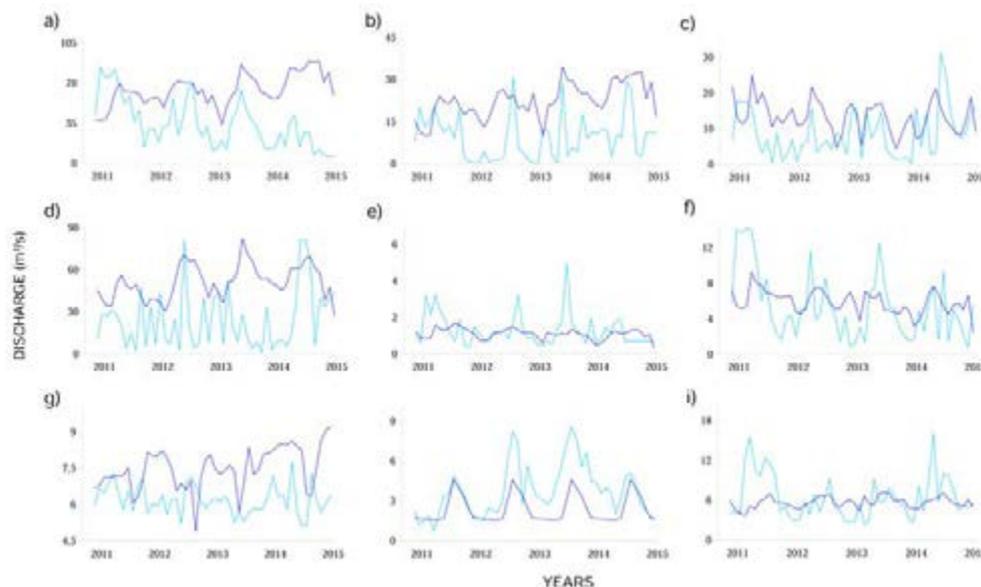


Figure 13

Observed and modeled flow time series using MLR and Multicollinearity Analysis (validation stage). a) Paute AJ Dudas, b) Gualaceo, c) Tomebamba, d) Paute DJ Gualaceo, e) Surucucho, f) Matadero, g) Collay, h) Dudas, i) Mazar

**Table 6**

Metrics results at the calibration and the validation stage

STATIONS	CALIBRATION STAGE				VALIDATION STAGE			
	RMSE	NSE	R2	KGE	RMSE	NSE	R2	KGE
PAUTE AJ DUDAS	26.2	0.32	0.32	0.38	39.73	-7.51	0.02	-0.36
GUALACEO	15.01	0.16	0.16	0.14	16.05	-5.86	0.02	-0.09
TOMBAMBA	11.46	0.17	0.17	0.18	8.76	-2.69	0.03	-0.01
PAUTE DJ GUALACEO	26.45	0.21	0.21	0.24	35	-7.26	0.02	-0.24
SURURCUCHO	0.72	0.18	0.18	0.19	0.89	-8.96	0.03	-1.25
MATADERO	4.03	0.08	0.08	-0.02	3.66	-6.92	0.07	-1.02
COLLAY	1.69	0.21	0.21	0.23	1.70	-2.83	0.01	-0.15
DUDAS	1.29	0.40	0.40	0.48	1.98	-2.77	0.45	-0.10
MAZAR	2.9	0.10	0.10	0.03	3.39	-15.73	0.02	-2.13

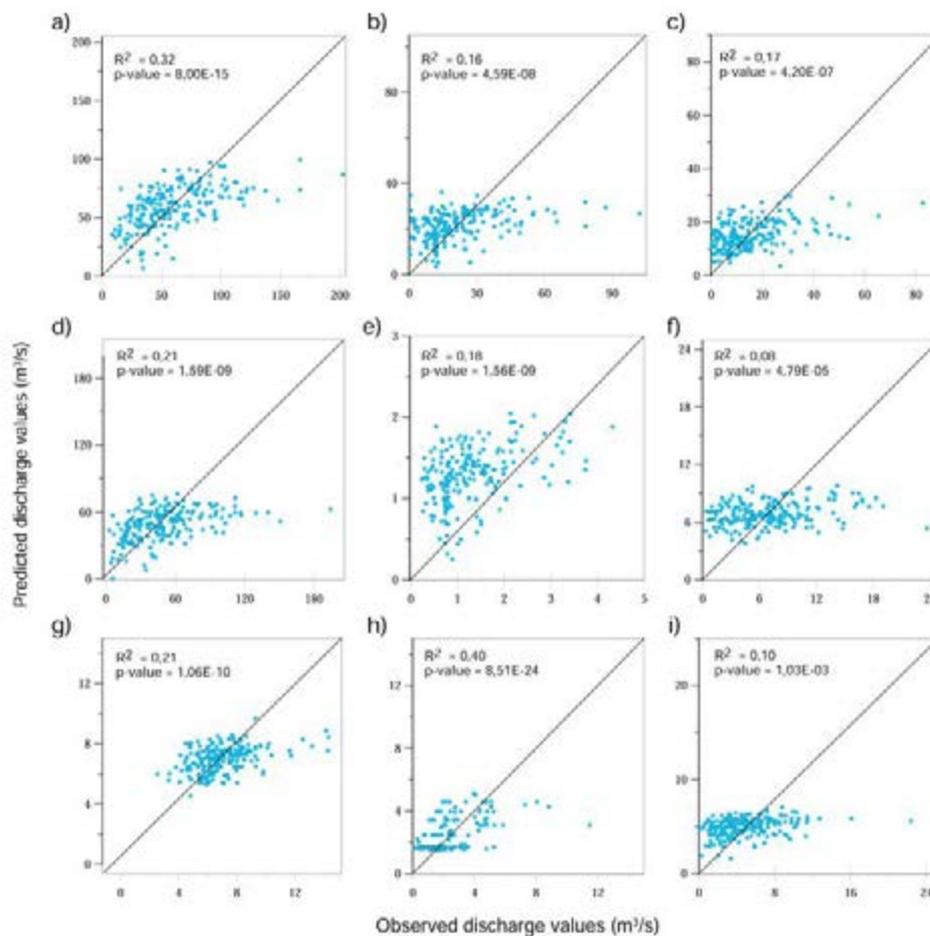
Using VIF to select the indices as variables in every model gave us different results for every station, and these results are not related to the functioning of indices but only to methodology. In the calibration stage, Dudas obtained the closest to a good result on correlation, NSE, and KGE, with 0.40 and 0.48, respectively. Paute AJ Dudas gets another close result to satisfactory with 0.38 in KGE. The worst fits occur on Matadero and Gualaceo, reaching even negative values on KGE (0.14 to -0.02). In this model, it is very clear that the results are not improved by using more variables. As a result, Dudas only uses one index to obtain better results in comparison with Mazar, which uses 5 (the maximum number of variables in each station) and gets poor results. The noise in this methodology does not influence it drastically since all RMSEs tend to maintain low values, which means the results are a show of the functioning itself. Dudas has one variable, Paute AJ Dudas 5, Matadero, and Gualaceo 1 and 2, proving

that only one index is enough to capture the variability of the discharges if it is related to the environment of the station.

In the validation stage, Tomebamba is the “best” fit model with -2.69 ; the rest have also negative values on NSE, in KGE. These results are similar, except in the case of Dudas with a 0.45 correlation, proving that is the station with the best results throughout the study. RMSE in both cases still obeys the initial values of the data to show results more than the model prediction, which is why the stations with the highest values, Paute AJ Dudas and Paute DJ Gualaceo, have high RSME, but their performance is very different, with Paute AJ Dudas the second-best station to fit the models.

Figure 14

Dispersion plots of observed and modeled discharge using MLR and Multicollinearity Analysis (calibration stage). a) Paute AJ Dudas, b) Gualaceo, c) Tomebamba, d) Paute DJ Gualaceo, e) Surucucho, f) Matadero, g) Collay, h) Dudas, i) Mazar



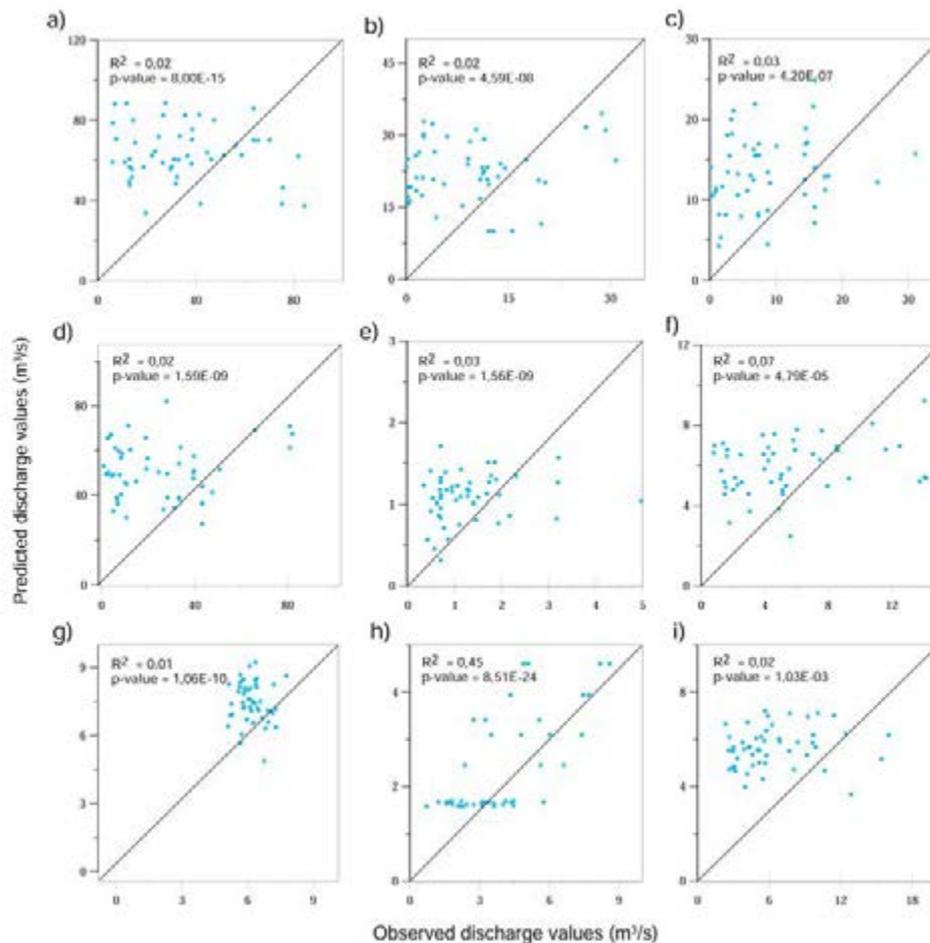
A study made in Indonesia that used SOI, Niño 3.4, and IOD indices to analyze the Java discharge regimen identified that the results on LRM with KGE metrics were best when the model included more variables (Nugroho, Tamagawa and Harada, 2022). This also happened in our study using this specific methodology, where we used a different number of variables for every station. In addition, from a general perspective, when using 20 components as variables on ACP also happened.

When we analyze Figure 14f, we can see the worst result: Matadero used only PWR to get the predictions, and we can see no relation between the behavior of the original series and

the predicted one. P- value maintains low in every station which means they are statistically representative. RSME is dominated by the same behavior as in the two other methodologies.

Figure 15

Dispersion plots of observed and modeled discharge using MLR and Multicollinearity Analysis (validation stage). a) Paute AJ Dudas, b) Gualaceo, c) Tomebamba, d) Paute DJ Gualaceo, e) Surucucho, f) Matadero, g) Collay, h) Dudas, i) Mazar



Even when the best models tend to have more variables, the variables change for every station; this is proven because stations with the same indices can have bad or good indices and are independent of each other. Also, there are stations that, using the same number of indices, can have better results than other stations. The results are directly linked with how the values of the time series couple with the indices data in an individual way. And the value of the data itself does not influence the results.

There are no similar indices used in all scenarios, but the ones that are present in the two methodologies (VIF and using monthly delays) are: NP, PWR, CAR, and CIP; also, this one has the most influence on loading factors in ACP (NP and PWR). The two better-fitting stations are Dudas and Paute AJ Dudas since this one does not have extreme events as frequently as the other stations, allowing LRM to work better. This methodology is good for approximation but it is limited to mathematical procedure because the only input data is the time series and not any environmental variable that can help to understand the hydrological behaviors (not environmental context or another variable). The stations with the worst results are Mazar, Gualaceo, and Matadero. Also, the validation range time could be amplified to improve the

models by catching more variability, especially in the extreme events that are not so frequent. In the Amazon Basin, research determined that ENSO domains the conditions of the discharges but does not fully understand the behavior of each variability mode and its effects. This is caused by the temporal SST anomalies, their magnitude, and their position in the equatorial Pacific Ocean. The study also points out that some statistical tests, such as Kendall, could not be suitable for certain regions. With a large-basin memory, it leads to autocorrelation and misleading significance (Towner et al., 2020). This is not only observed in our study but also evidenced in each of the methodologies used.

Conclusions

The best models can follow the original data tendency but cannot predict whether the values are too low or too high. The best scenarios were obtained when using PCA as a general view, but the best values in metrics were obtained when using a 1-month delay. The models except Dudas and Paute AJ Dudas mostly didn't reach a satisfactory result in all metrics (NSE, KGE, R2). This means that indices are not enough as discharge predictors. RSME is not a good metric in our study since its values were associated with the data and not with the model's performance in all cases. It shows a very light tendency on PCA, lowering their values. The validation stage determines that the models do not have robustness, and this is directly related to the initial conditions of our models not applying to different contexts. We have several indices associated with ENSO that do not represent any relevance in the study (AO, BEST, ENSO, GIAM, MEI V2, NAO, Niño 3, Niño 3.4, ONI, SOI, TPI, and IPO). It is important to recognize if this index has a direct or indirect effect on the zone and which of his variability modes better represents it. Finally, it is important to highlight that being a first approximation to the use of all teleconnection indices to measure their predictive capacity in the Paute river basin, the results are very interpretable. It is essential to mention that the teleconnection indices (signals and modes of variability), in addition to providing global information, maintain a well-studied atmospheric and oceanic circulation. This makes it possible not only to observe climate variability on a large spatial and temporal scale but also to analyze variables dependent on these phenomena without having to resort to breaking them down for a single study.

Acknowledgments

The authors would like to thank the INAMHI for the information provided. This work was founded for Corporación Ecuatoriana para el Desarrollo de la Investigación y la Academia (CEDIA) within the research project “Análisis Nexus agua-alimentos-energía-servicios ecosistémicos ante cambios del clima, uso del suelo y población. Un enfoque novedoso para el desarrollo sostenible local a escala de una cuenca hidrográfica” and Vicerrectorado de Investigaciones de la Universidad de Cuenca (VIUC).

References

- Bazo, J., Lorenzo, M. D. L. N., & Porfirio Da Rocha, R. (2013). Relationship between monthly rainfall in NW Peru and tropical sea surface temperature. *Advances in Meteorology*, 2013. <https://doi.org/10.1155/2013/152875>
- Cai, W., McPhaden, M. J., Grimm, A. M., Rodrigues, R. R., Taschetto, A. S., Garreaud, R. D., Dewitte, B., Poveda, G., Ham, Y.-G., Santoso, A., Ng, B., Anderson, W., Wang, G., Geng, T., Jo, H.-S., Marengo, J. A., Alves, L. M., Osman, M., Li, S., ... Vera, C. (2020). Climate impacts of the El Niño–Southern Oscillation on South America. *Nature Reviews Earth & Environment*, 1(4), 215–231. <https://doi.org/10.1038/s43017-020-0040-3>

- Campozano, L., Célleri, R., Trachte, K., Bendix, J., & Samaniego, E. (2016). Rainfall and Cloud Dynamics in the Andes: A Southern Ecuador Case Study. *Advances in Meteorology*, 2016, 1–15. <https://doi.org/10.1155/2016/3192765>
- Carreric, A. (2019). *Enso diversity and global warming* (Doctoral dissertation, Université Paul Sabatier-Toulouse III).
- CELEC EP. (2013). *Actualización Del Estudio De Impacto Ambiental Y Plan De Manejo Ex Post De La Central Paute Molino (Update of Environmental Impact Assessment and Post-Management Plan of Paute Molino Power Plant)*.
- Celleri, R., Willems, P., Buytaert, W., & Feyen, J. (2007). Space–time rainfall variability in the Paute basin, Ecuadorian Andes. *Hydrological Processes*, 21(24), 3316–3327. <https://doi.org/10.1002/hyp.6575>
- Choubin, B., Khalighi-Sigaroodi, S., Malekian, A., & Kişi, Ö. (2016). Multiple linear regression, multi-layer perceptron network and adaptive neuro-fuzzy inference system for forecasting precipitation based on large-scale climate signals. *Hydrological Sciences Journal*, 61(6), 1001–1009. <https://doi.org/10.1080/02626667.2014.966721>
- Córdoba Machado, S., Palomino Lemus, R., Gámiz Fortis, S. R., Castro Díez, Y., & EstebanParra, M. J. (2015). Assessing the impact of El Niño Modoki on seasonal precipitation in Colombia. *Global and Planetary Change*, 124, 41–61. <https://doi.org/10.1016/j.gloplacha.2014.11.003>
- Crimmins, M. (2014). Southwestern Monsoon. Climate Assessment for the SouthWest. <https://climas.arizona.edu/sw-climate/monsoon>
- Dima, M., & Lohmann, G. (2004). Fundamental and derived modes of climate variability: concept and application to interannual time-scales. *Tellus A: Dynamic Meteorology and Oceanography*, 56(3), 229. <https://doi.org/10.3402/tellusa.v56i3.14415>
- Esha, R. I., Imteaz, M. A., & Nazari, A. (2019). Assessing Gene Expression Programming as a technique for seasonal streamflow prediction: A case study of NSW. *IOP Conference Series: Earth and Environmental Science*, 351(1), 012004. <https://doi.org/10.1088/1755-1315/351/1/012004>
- Espino Sánchez, M. A. (2014). *Patrones de variabilidad ambiental y las pesquerías en el Pacífico Sud Este (Patterns of Environmental Variability and Fisheries in the Southeast Pacific)*. Universidad Nacional Mayor de San Marcos.
- Fan, M., & Schneider, E. K. (2012). Observed Decadal North Atlantic Tripole SST Variability. Part I: Weather Noise Forcing and Coupled Response. *Journal of the Atmospheric Sciences*, 69(1), 35–50. <https://doi.org/10.1175/JAS-D-11-018.1>
- Giddings, L., & Soto, M. (2006). *Teleconexiones y precipitación en América del Sur [Teleconnections and precipitation in South America]*. 6, 13–20.
- Gobena, A. K., Weber, F. A., & Fleming, S. W. (2013). The Role of Large-Scale Climate Modes in Regional Streamflow Variability and Implications for Water Supply Forecasting: A Case Study of the Canadian Columbia River Basin. *Atmosphere-Ocean*, 51(4), 380–391. <https://doi.org/10.1080/07055900.2012.759899>
- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2021). *An Introduction to Statistical Learning* (2nd ed.). Springer US. <https://doi.org/10.1007/978-1-0716-1418-1>
- Hatzaki, M., Flocas, H. A., Asimakopoulos, D. N., & Maheras, P. (2007). The eastern Mediterranean teleconnection pattern: identification and definition. *International Journal of Climatology*, 27(6), 727–737. <https://doi.org/10.1002/joc.1429>
- IDEAM - UNAL (2018). *Variabilidad Climática y el cambio climático en Colombia (1era ed.) [Climate Variability and Climate Change in Colombia (1st ed.)]*. IDEAM

- Contreras, J., Ballari, D., & Samaniego, E. (2017). EJE 02-09 Optimización de una red de monitoreo de precipitación usando modelos Geoestadísticos: caso de estudio en la cuenca del río Paute, Ecuador [AXIS 02-09 Optimization of a precipitation monitoring network using geostatistical models: case study in the Paute river basin, Ecuador]. *Memorias Y Boletines De La Universidad Del Azuay*, 1(XVI), 115–124. <https://doi.org/10.33324/memorias.v1iXVI.55>
- Knoben, W. J. M., Freer, J. E., & Woods, R. A. (2019). Technical note: Inherent benchmark or not? Comparing Nash–Sutcliffe and Kling–Gupta efficiency scores. *Hydrology and Earth System Sciences*, 23(10), 4323–4331. <https://doi.org/10.5194/hess-23-4323-2019>
- Krause, P., Boyle, D. P., & Bäse, F. (2005). Comparison of different efficiency criteria for hydrological model assessment. *Advances in Geosciences*, 5, 89–97. <https://doi.org/10.5194/adgeo-5-89-2005>
- Kundzewicz, Szwed, & Pińskwar. (2019). Climate Variability and Floods—A global Review. *Water*, 11(7), 1399. <https://doi.org/10.3390/w11071399>
- Matute, V. (2014). *Análisis De Factibilidad De Generación Eléctrica A Pie De La Presa De Chanlud* (Doctoral dissertation, Universidad de Cuenca).
- McGregor, G. (2017). Hydroclimatology, modes of climatic variability and stream flow, lake and groundwater level variability. *Progress in Physical Geography: Earth and Environment*, 41(4), 496–512. <https://doi.org/10.1177/0309133317726537>
- Meyer, T. (2010). Technical note: Root Mean Square Error Compared to, and Contrasted with, Standard Deviation. *Surveying and Land Information Science*, 72(3), 107-108
- Morán-Tejeda, E., Bazo, J., López-Moreno, J. I., Aguilar, E., Azorín-Molina, C., Sanchez-Lorenzo, A., Martínez, R., Nieto, J. J., Mejía, R., Martín-Hernández, N., & Vicente-Serrano, S. M. (2016). Climate trends and variability in Ecuador (1966-2011). *International Journal of Climatology*, 36(11), 3839–3855. <https://doi.org/10.1002/joc.4597>
- Moriassi, D., Arnold, J., Van Liew, M., Bingner, R., Harmel, R., & Veith, T. (2007). Model Evaluation Guidelines for Systematic Quantification of Accuracy in Watershed Simulations. *Transactions of the ASABE*, 50(3), 885–900. <https://doi.org/10.13031/2013.23153>
- Nugroho, A. R., Tamagawa, I., & Harada, M. (2022). Spatiotemporal Analysis on the Teleconnection of ENSO and IOD to the Stream Flow Regimes in Java, Indonesia. *Water (Switzerland)*, 14(2). <https://doi.org/10.3390/w14020168>
- Orbes, J., & Peralta, T. (2017). *Estado del arte en Manejo de Sedimentos en cuencas Andinas en el Ecuador, caso de estudio: cuenca del Río Paute [State of the Art in Sediment Management in Andean Watersheds in Ecuador; Case Study: Paute River Basin]*. (Bachelor dissertation, Universidad de Cuenca)
- Fontaine, G., Narváez, I., and Cisneros, P. (2008). *[Geo Ecuador 2008: State of the Environment Report]*. FLACSO. Quito, Ecuador. Available online at: <https://biblio.flacsoandes.edu.ec/libros/digital/41444.pdf> (accessed April 24, 2023).
- Rea, A., & Rea, W. (2016). How Many Components should be Retained from a Multivariate Time Series PCA ?
- Schneider, T., Hampel, H., Mosquera, P. V., Tylmann, W., & Grosjean, M. (2018). Paleo-ENSO revisited: Ecuadorian Lake Pallcacocha does not reveal a conclusive El Niño signal. *Global and Planetary Change*, 168, 54–66. <https://doi.org/10.1016/j.gloplacha.2018.06.004>
- Schöngart, J., Junk, W. J., Piedade, M. T. F., Ayres, J. M., Hüttermann, A., & Worbes, M. (2004). Teleconnection between tree growth in the Amazonian floodplains and the El Niño–Southern Oscillation effect. *Global Change Biology*, 10(5), 683–692. <https://doi.org/10.1111/j.1529-8817.2003.00754.x>
- Sedano, R. (2017). *Influencia de la variabilidad climática y factores antrópicos en los extremos hidrológicos en el Valle Alto del río Cauca, Colombia [Influence of Climate Variability and Anthropogenic Factors on Hydrological Extremes in the Upper Cauca River Valley, Colombia]*. (Doctoral dissertation, Universitat Politècnica de Valencia).

- Shuhaida, I., & Shabri, A. (2014). Stream flow forecasting using principal component analysis and least square support vector machine. *Journal of Applied Science and Agriculture*, 9(11), 170–180.
- Sotomayor, G., Hampel, H., & Vázquez, R. F. (2018). Water quality assessment with emphasis in parameter optimisation using pattern recognition methods and genetic algorithm. *Water Research*, 130, 353–362. <https://doi.org/10.1016/j.watres.2017.12.010>
- Towner, J., Cloke, H. L., Lavado, W., Santini, W., Bazo, J., Coughlan de Perez, E., & Stephens, E. M. (2020). Attribution of Amazon floods to modes of climate variability: A review. *Meteorological Applications*, 27(5). <https://doi.org/10.1002/met.1949>
- Vega Vilca, J. C., & Guzman, J. (2011). Regresion PLS y PCA Como Solución al Problema de Multicolinealidad en Regresion Multiple. *Revista de Matemática: Teoría y Aplicaciones*, 18(1), 9. <https://doi.org/10.15517/rmta.v18i1.2111>
- Wang, C., & Enfield, D. B. (2001). The Tropical Western Hemisphere Warm Pool. *Geophysical Research Letters*, 28(8), 1635–1638. <https://doi.org/10.1029/2000GL011763>
- Ward, E., Buytaert, W., Peaver, L., & Wheeler, H. (2011). Evaluation of precipitation products over complex mountainous terrain: A water resources perspective. *Advances in Water Resources*, 34(10), 1222–1231. <https://doi.org/10.1016/j.advwatres.2011.05.007>
- Whitfield, P. H., Moore, R. D. (Dan), Fleming, S. W., & Zawadzki, A. (2010). Pacific Decadal Oscillation and the Hydroclimatology of Western Canada—Review and Prospects. *Canadian Water Resources Journal*, 35(1), 1–28. <https://doi.org/10.4296/cwrj3501001>

Análisis de la Calidad de las Asociaciones de Contaminantes y Variables Meteorológicas por Estacionalidad

Analysis of the Quality of the Associations of Pollutants and Meteorological Variables by Seasonality

Jorge Luis Zambrano-Martinez¹ <https://orcid.org/0000-0002-5339-7860>, Marcos Orellana¹ <https://orcid.org/0000-0002-3671-9362>, Agustín Ferrari¹ <https://orcid.org/0009-0009-5696-8954>, Alex Coro¹ <https://orcid.org/0009-0009-6020-3751>, Sebastian Heras¹ <https://orcid.org/0009-0001-9148-0139>

¹Laboratorio de Investigación y Desarrollo en Informática (LIDI)
Universidad del Azuay, Cuenca, Ecuador
jorge.zambrano@uazuay.edu.ec, marore@uazuay.edu.ec,
daferrarif99@es.uazuay.edu.ec,
alex.coro@es.uazuay.edu.ec,
sebastianheras28@es.uazuay.edu.ec



Esta obra está bajo una licencia internacional
Creative Commons Atribución-NoComercial 4.0

Enviado: 2023/07/14

Aceptado: 2023/07/29

Publicado: 2023/10/15

Resumen

Para lograr una evaluación más precisa de la calidad del aire, es necesario conocer la relación que existe entre las variables meteorológicas y los distintos contaminantes atmosféricos; lo que también tendrá la finalidad de evitar los riesgos presentes tanto en el ecosistema como en la salud de los humanos en un futuro próximo. El problema radica en encontrar una asociación entre los contaminantes atmosféricos y las variables meteorológicas, que existe en los modelos y métodos de categorización que pueden ser empleados. Debido a esta razón, el objetivo de este artículo es analizar la calidad de las asociaciones de contaminantes y variables meteorológicas por estacionalidad utilizando árboles de decisión para encontrar conocimiento que permita localizar patrones que serán importantes para el análisis ambiental. Por consiguiente, al aplicar árboles de decisión se pudo lograr un control periódico de la calidad de las asociaciones de contaminantes y variables meteorológicas, cuya validación del nivel de confianza de las reglas de asociación es mayor al 70% en los meses estudiados.

Sumario: Introducción, Trabajos Relacionados, Metodología, Resultados y Conclusiones.

Como citar: Zambrano-Martinez, J. L., Orellana, M., Ferrari, A., Coro, A. & Heras, S. (2023). Análisis de la Calidad de las Asociaciones de Contaminantes y Variables Meteorológicas por Estacionalidad. *Revista Tecnológica - Espol*, 35(2), 50-60.

Recuperado a partir de <http://www.rte.espol.edu.ec/index.php/tecnologica/article/view/1052>

Palabras clave: Árboles de decisión, discretización, inteligencia artificial, minería de datos, reglas de asociación.

Abstract

To achieve a more accurate assessment of air quality, it is necessary to know the relationship between meteorological variables and the different air pollutants; this will also be aimed at avoiding the risks present both in the ecosystem and in human health in the future. The problem begins with finding an association between air pollutants and meteorological variables in the models and categorization methods that can be used. Due to this reason, the objective of this article is to analyze the quality of the associations between pollutants and meteorological variables by seasonality using decision trees to find knowledge that allows locating patterns that will be important for environmental analysis. Therefore, it was possible to achieve a periodic control of the quality of the associations of pollutants and meteorological variables, whose validation of the confidence level of the association rules is greater than 70% in the months studied.

Keywords: Decision trees, discretization, artificial intelligence, data mining, association rules.

Introducción

Actualmente, la contaminación atmosférica es uno de los problemas ambientales más graves del mundo, debido a la gran cantidad de contaminantes emitidos por los humanos, entre ellos destaca el material particulado, menor a 2.5 micrómetros ($PM_{2.5}$); el material particulado, menor a 10 micrómetros (PM_{10}); el dióxido de azufre (SO_2); el monóxido de carbono (CO); el ozono (O_3); y el dióxido de nitrógeno (NO_2). Estos contaminantes presentan un efecto dañino a la salud debido a que provocan diferentes afecciones pulmonares, cerebrales y cardiovasculares, lo que reduce la esperanza de vida (Manisalidis et al., 2020). En las zonas rurales persiste un menor índice de contaminación en comparación con las zonas urbanas, debido a la alta densidad demográfica, conjuntamente con el desarrollo de la industria y el transporte. Las consecuencias que conlleva están relacionadas con problemas de salud, que traen consigo preocupación en los gobiernos. Así, los gobernantes tienen que enfocarse en una adecuada gestión de la calidad del aire, utilizando diferentes estrategias y herramientas que permitan una correcta interpretación de los datos obtenidos en la atmósfera.

Para el análisis de este tipo de estudios es necesaria la recolección de los datos provenientes de las estaciones de monitoreo meteorológico continuo, con la finalidad de conocer la calidad del aire y a su vez encontrar los problemas de contaminación atmosférica en una zona. Además, la relación de los contaminantes atmosféricos con las variables meteorológicas, como la humedad, temperatura, velocidad del aire, entre otros, tienden a afectar el clima del entorno, y dan como resultado el deterioro de la salud de las personas (Whiteman et al., 2014). Del mismo modo, Ahmed et al. (2020) sostienen que las concentraciones de contaminación se enfocan en zonas específicas de interés o parques industriales. Huang et al. (2018) determinó que, si no se toman las debidas precauciones, como el cambio de combustibles a más limpios o la depuración de buses anticuados, la contaminación atmosférica continuará en aumento de manera considerable.

Siendo así, los gobiernos locales se encuentran recolectando de manera constante los datos de las variables meteorológicas y contaminantes atmosféricos, a través de estaciones meteorológicas equipadas con sensores, que recopilan toda la información del ambiente para

luego ser procesada y brindar conocimiento. Se utilizan técnicas de minería de datos para el tratamiento de la información obtenida y el descubrimiento de patrones de comportamiento como árboles de decisión, regresiones logísticas, algoritmos genéticos, entre otros.

En este estudio, para realizar el análisis de los patrones de comportamiento y crear modelos de predicción con los datos de los contaminantes atmosféricos y variables meteorológicas en la ciudad de Cuenca – Ecuador, se utilizan los árboles de decisión con la finalidad de validar el grado de confianza de las reglas de asociación obtenidas en un previo estudio (Orellana et al., 2021).

Este artículo está estructurado de la siguiente manera: La Sección 2 presenta los trabajos relacionados con métodos semejantes; la Sección 3 expone la metodología utilizada para llevar a cabo esta investigación; la Sección 4 explica los resultados que se han obtenido tras la aplicación de la metodología; y la Sección 5 presenta las conclusiones obtenidas y los trabajos futuros.

Trabajos Relacionados

Existen investigaciones que estudian a las variables meteorológicas y a los contaminantes atmosféricos y permiten obtener información para distintos propósitos, como predecir nuevo conocimiento para determinar el nivel de calidad de la atmósfera y su posible gestión. A continuación, se exponen las investigaciones que están vinculadas a técnicas de minería de datos que extraen los conocimientos más importantes en este campo.

En el estudio realizado por García et al. (2020) se utilizan métodos de minería de datos para cuantificar el impacto ambiental producido por una empresa Courier en Lima y Callao, en Perú. Como punto inicial, los autores aplican métodos de clasificación a las variables y clusterización (OneRule y K-means), lo cual permite detectar los de mayor demanda de despachos (4 grupos) y determinar el nivel de emisiones de KgCO_2 diarias, separadas por una temporalidad mensual y por tipo de vehículo utilizado. En cuanto a la investigación expuesta por Gayathri et al. (2020), propone la utilización de un modelo basado en técnicas de minería de datos para la predicción de la contaminación del aire, específicamente la aplicación de algoritmo de árbol de decisión *C4.5*. Los autores descomponen dicho modelo en cinco etapas: recolección de datos, preprocesamiento de los datos, árbol de decisión, datos de prueba, predicción, con lo cual concluyen que el sistema propuesto ayudará a mejorar la predicción.

Acorde a Siwek et al. (2016), se realizan los modelos de predicción a diferentes contaminantes como PM_{10} , $\text{PM}_{2.5}$, NO_2 , SO_2 y O_3 . Sin embargo, uno de los contaminantes más nocivos para la salud humana es $\text{PM}_{2.5}$, debido a la acumulación en el sistema respiratorio, disminuyendo el correcto funcionamiento pulmonar y aumentando las enfermedades respiratorias. El método que utilizan para predecir el valor de $\text{PM}_{2.5}$ es el de bosques aleatorios y los resultados demuestran que el modelo presentado es viable para aplicarlo en cualquier ciudad contaminada. Es importante considerar el índice de calidad del aire (AQI) utilizado por los gobiernos para determinar el nivel de calidad que se presenta en la atmósfera relacionados con los contaminantes. Asimismo, el trabajo realizado por Huang et al. (2018) utiliza algoritmos de redes neuronales, basados en el índice $\text{PM}_{2.5}$, para llegar a pronosticar la contaminación del ambiente. La conclusión obtenida en este estudio es que la única variable atmosférica vinculada con el contaminante $\text{PM}_{2.5}$ es la humedad, debido a que ayuda a la disipación del contaminante.

Otro contaminante altamente perjudicial es el PM_{10} , de acuerdo con Althuwaynee et al. (2020), debido a que analizan las correlaciones que existen entre PM_{10} y otros contaminantes

como el O_3 y SO_2 . Los datos son obtenidos en Kuala Lumpur, Malasia, y en su análisis se observa que existe una relación lineal directa y positiva entre PM_{10} y SO_2 , mientras que existe una relación semi lineal entre PM_{10} y O_3 .

Continuando con las investigaciones relacionadas con redes neuronales, está la realizada por Athira et al. (2018), que utilizó el conjunto de datos de AirNet, de pronóstico de contaminación del aire de cinco días del Centro Nacional de Monitoreo Ambiental de China (CNEMC) y datos meteorológicos del sistema global de predicción (GFS), donde se aplican modelos de aprendizaje profundo, como red neuronal recurrente (RNN), redes de gran memoria a corto plazo (LSTM) y unidades recurrentes cerradas (GRU). En otra investigación, realizada por Birant (2011), se presentan los resultados de comparación de distintos algoritmos de árboles de decisión como *C4.5*, *CART*, *NBTree*, *BFTree*, *LADTree*, *REPTree*, árbol aleatorio, bosques aleatorios, y árbol modelo logístico (LMT), para clasificar y predecir los niveles de emisión del SO_x , utilizando los datos recopilados de alrededor de 800 instalaciones industriales en Izmir, Turquía. Los autores compararon individualmente cada nivel de emisión del SO_x (bajo, medio, alto y muy alto), obteniendo los mejores resultados en la categoría de bajo nivel. Así, se pudo establecer que los niveles de emisión pueden clasificarse y predecirse con éxito en el 82.4% de los casos.

Por otro lado, Martínez-España et al. (2018), analizaron diferentes técnicas de aprendizaje automático para predecir los niveles del O_3 en la región de Murcia, España. Las técnicas analizadas son de bosques aleatorios, árboles de decisión y K vecinos más próximos (*KNN*); la técnica de bosques aleatorios fue la que más se ajustó. En los resultados obtenidos se puede señalar que entre los parámetros que más influyen en la predicción del ozono se encuentran variables climáticas relacionadas con la temperatura, la humedad y el viento.

Así pues, en este estudio se realiza un análisis de los patrones de comportamiento y la creación de modelos de predicción con los datos obtenidos de los contaminantes atmosféricos y variables meteorológicas en la ciudad de Cuenca - Ecuador, a través de árboles de decisión para ratificar el nivel de confianza de las reglas de asociación previamente establecidas en trabajos previos.

Metodología

En la presente investigación, la metodología que se utilizó para llevar a cabo la experimentación se denomina ADDIE (Análisis, Diseño, Desarrollo, Implementación y Evaluación) y se encuentra representada en la Figura 1. Este modelo es considerado uno de los más utilizados y sus cinco fases comprenden un camino o proceso de aprendizaje.

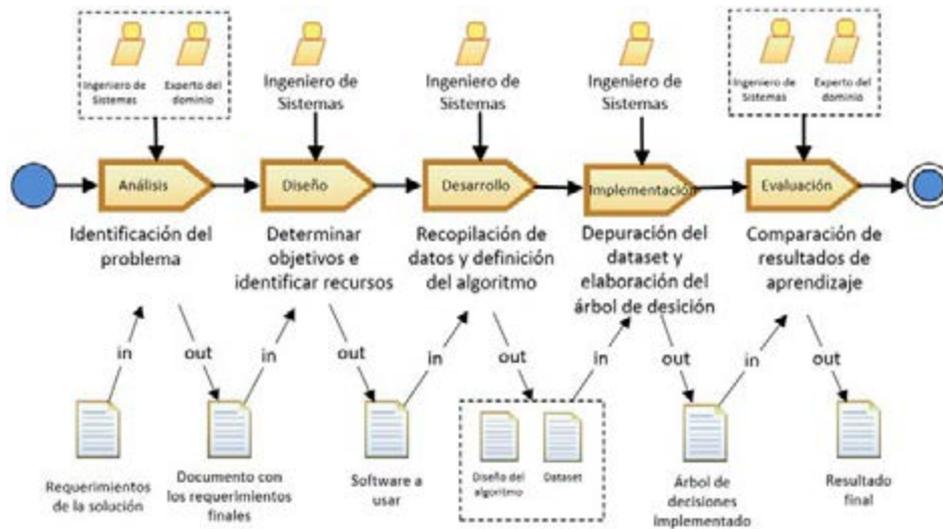
A continuación, se presentan, con más detalle, las fases de la metodología ADDIE aplicadas a esta investigación:

- **Análisis:** La propuesta de esta investigación es realizar un análisis de la calidad de las asociaciones de contaminantes y variables meteorológicas por estacionalidad, aplicando árboles de decisión a fin de verificar si las reglas de asociación expuestas en la investigación de Orellana et al. (2021) se cumplen con su calidad.
- **Diseño:** Considerando el análisis de la calidad de las asociaciones de contaminantes y variables meteorológicas, el principal objetivo de este análisis es el cumplimiento de la calidad de las reglas de asociación y así definir el comportamiento de las reglas de asociación correspondientes a los meses específicos. Es por esa razón que, con

este objetivo en mente, se ha definido trabajar con la herramienta de minería de datos aplicando árboles de decisión, que es una técnica de inteligencia artificial utilizando el software *RapidMiner*. Esta herramienta reduce el uso del código para el modelado de los datos, agilizando el análisis y el procesamiento de grandes cantidades de datos.

- **Desarrollo:** Para la selección de los parámetros necesarios, se procede en función de la recopilación de los datos para definir un algoritmo con la mayor precisión posible.
- **Implementación:** Para realizar la implementación, se obtuvo un conjunto de datos de los contaminantes atmosféricos tomados entre el mes de enero a diciembre del año 2018 en la ciudad de Cuenca, Ecuador. Para realizar el preprocesamiento de los datos, donde se realiza la depuración y limpieza, se utilizó una plataforma de ciencia de datos denominada *RapidMiner*; luego se procedió a realizar las pruebas correspondientes.
- **Evaluación:** Mediante el ingreso de los datos al software mencionado se determinaron los parámetros más importantes del conjunto de datos, para generar un árbol de decisión correspondiente a cada mes y verificar que las reglas de asociación obtenidas se cumplan o no en ciertos meses.

Figura 1
Diseño de la metodología ADDIE



Preprocesamiento de datos

Como se conoce, al recolectar los datos por medio de los sensores de las estaciones meteorológicas, estos pueden presentar anomalías o valores atípicos, ya sea por diversas razones como interrupciones en la fuente energética o errores de calibración en los sensores que están utilizando. Por lo tanto, la recolección de datos de los contaminantes atmosféricos es tomada cada diez minutos y las variables meteorológicas con una frecuencia de cada minuto. La recolección de los datos entre los meses de enero a diciembre del 2018 fue de 52,559 registros.

Este conjunto de datos contiene atributos que son dependientes de otros, en otras palabras, los valores finales de una variable dependen de la mezcla de cualidades que sean partícipes en una determinada dimensión del estudio.

Cabe recalcar que se implementó la depuración de valores no encontrados, con la finalidad de descartar los datos que contienen valores vacíos o nulos. Luego de realizar el preprocesamiento de los datos, se obtuvieron un total de 39,507 registros y se eliminaron un total de 8,497 datos, para obtener valores consistentes de comportamiento.

Aplicación de técnicas de discretización

La discretización es el procedimiento en el cual se divide el rango del atributo continuo en intervalos. Cada uno de los intervalos se etiqueta como un valor discreto, para luego a los datos originales asignarlos a los valores discretos (Hemada & Vijaya Lakshmi, 2013).

Este proceso tiene un enfoque importante en el preprocesamiento de datos para su utilización en técnicas de inteligencia artificial, ya que con la aplicación de un método eficaz de discretización no solo se puede reducir la demanda de memoria del sistema y mejorar la eficiencia de la minería de datos, sino que también hace que el conocimiento extraído del conjunto de datos sea más sólido y fácil de entender (Hemada & Vijaya Lakshmi, 2013).

Para este estudio, es necesario aplicar la discretización con cantidades similares de datos en cada grupo para utilizar las reglas de asociación presentadas por Orellana et al. (2021) y así analizarlas mediante árboles de decisión para obtener sus respectivas medidas de rendimiento.

Aplicación de árboles de decisión

El algoritmo de árboles de decisión es un modelo lógico parecido a un árbol binario construido con base en un conjunto de datos de entrenamiento. Este ayuda a predecir el valor de una variable objetivo mediante el uso de variables predictivas. Este es el método ideal para la clasificación de información y su posterior evaluación en sus diferentes escenarios (Alsagheer et al., 2017).

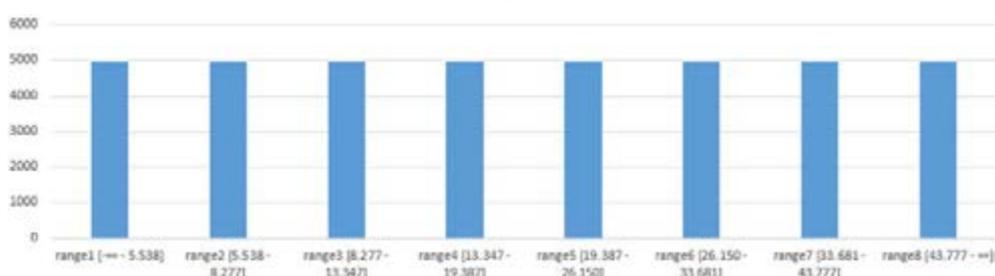
Resultados

Al momento de tener un conjunto de datos amplio, lo óptimo y recomendable es agrupar los datos que representan una misma unidad. Por lo tanto, se debe tomar en cuenta que la técnica sea fiable por el hecho de obtener patrones para observar con más detalle el comportamiento de los datos.

Técnica de discretización

Para tener un mejor enfoque en la investigación, se decidió crear varios conjuntos con una cantidad de datos equilibrada, como se puede observar en la Figura 2. A través de esta técnica de asociación, presentada por Orellana et al. (2021), se crearon ocho conjuntos con distintos rangos de datos para finalmente verificar las reglas de asociación.

Figura 2
Discretización por frecuencia

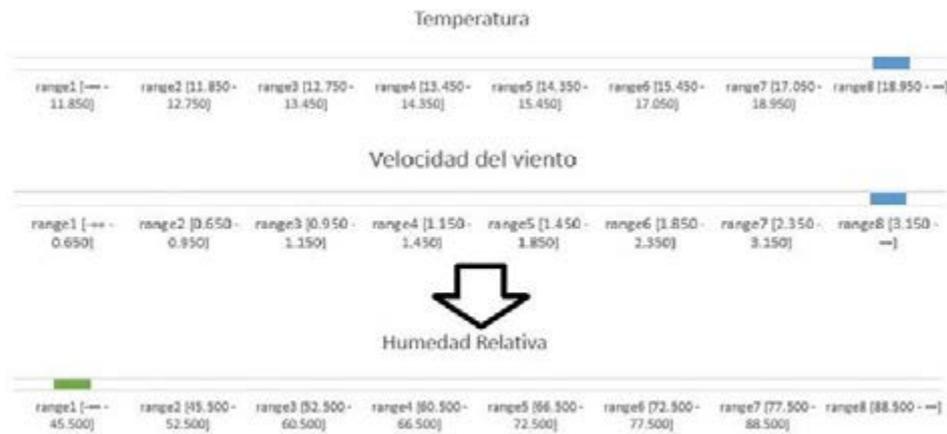


Reglas de asociación

Para este estudio se contó con tres reglas de asociación previamente definidas por Orellana et al. (2021) y mediante la discretización se obtuvieron las mismas reglas con el conjunto de datos.

La primera regla de asociación establece que al tener una temperatura que se encuentra en un rango superior a 18,950 °C, junto a un valor de la velocidad del viento superior a 3,150 m/h, produce como resultado una humedad relativa baja; en otras palabras, los datos de la humedad relativa se ubican menor igual a 45,5%. En la Figura 3, se puede observar la primera regla de asociación.

Figura 3
Primera regla de asociación



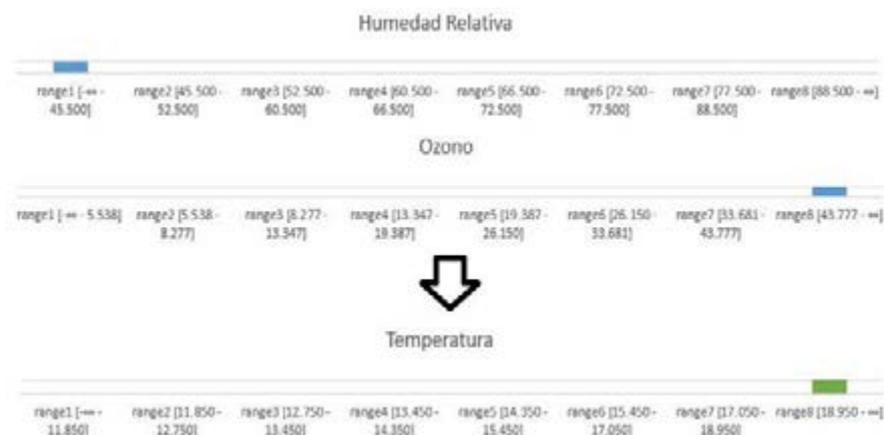
Por otro lado, la segunda regla de asociación propone que al presentarse una medición baja de O₃, menor a 5,538 µg/m³ y una humedad relativa alta, mayor a 88,5%, da como resultado valores altos del punto de rocío. Esto se puede apreciar en la Figura 4.

Figura 4
Segunda regla de asociación



Por último, la tercera regla de asociación plantea que, al presentarse una humedad relativa baja, menor a 45,5%, y una medida del O₃ alta, que supere el valor de 45 µg/m³, el resultado que se presenta es una temperatura alta, que puede sobrepasar los 18,950 °C. como se puede observar en la Figura 5.

Figura 5
Tercera regla de asociación



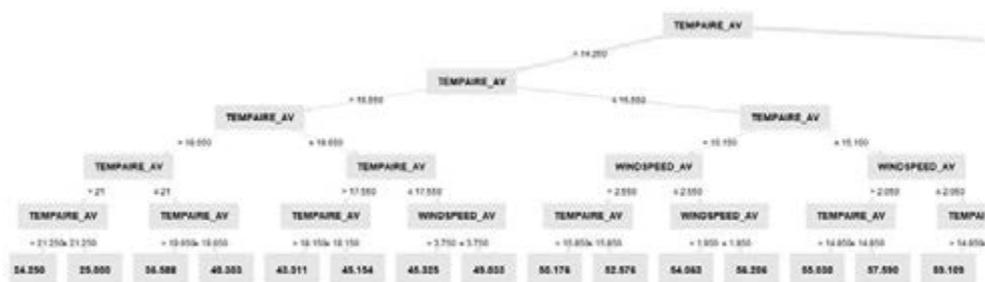
Árboles de decisión

La finalidad de aplicar los árboles de decisión es validar el nivel de confianza de las reglas de asociación presentadas en la subsección anterior. Por lo tanto, se realizó la validación del nivel de confianza con árboles de decisión mensualmente, desde enero hasta diciembre de 2018, y se observó la frecuencia con la que se cumplen dichas reglas de asociación.

Se aplicó este método a la primera regla y se obtuvo lo siguiente:

- La primera regla no se cumplió en 3 de los 12 meses analizados, por lo que representa el 25%, siendo los meses de junio, julio y agosto.
- Existieron dos meses en los que se pudo observar que la velocidad del viento no influyó para representar los niveles bajos propuestos de humedad relativa como se puede observar en la Figura 6.

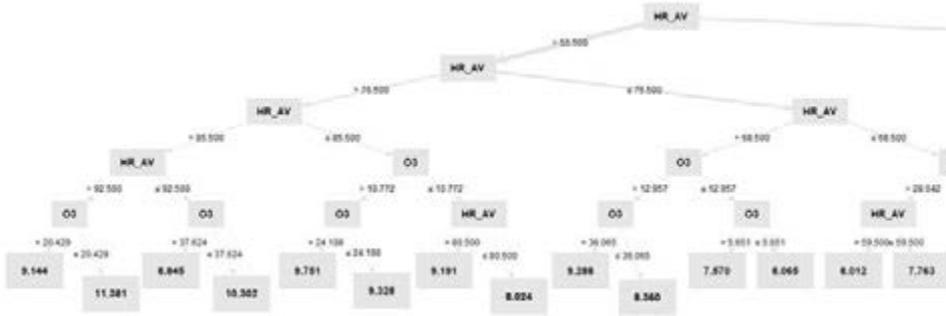
Figura 6
Árbol de decisión aplicado en la primera regla de asociación



En cuanto a la segunda regla de asociación, los resultados obtenidos fueron:

- Su medida de rendimiento fue del 66.7%; es decir, la regla llegó a cumplirse en ocho de los 12 meses. Sin embargo, cabe recalcar que en algunos casos la medición obtenida de O_3 es demasiado elevada, por lo que no se puede llegar a establecer con claridad lo que sucedió en el mes de enero.
- También existe el caso en el que los valores de O_3 llegan a ser los esperados pero la humedad relativa es baja y no llega a superar el 88.5%, lo cual influyó en la obtención de la medida de rendimiento de la regla, como se puede observar en la Figura 7.

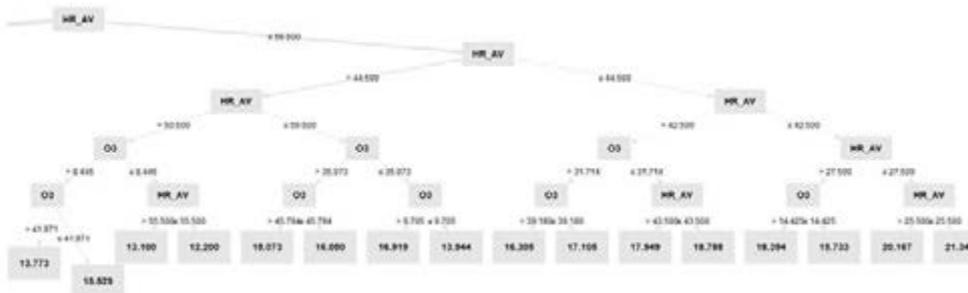
Figura 7
Árbol de decisión aplicado en la segunda regla de asociación



Por último, los resultados obtenidos en la tercera regla de asociación, como se puede observar en la Figura 8, fueron:

- En esta última regla se trabaja con O₃, humedad relativa y temperatura, cumpliéndose en el 75% de los meses excepto los meses de enero, julio y octubre.
- Igual existen casos problemáticos con el contaminante O₃, debido a que los valores medidos tienden a ser bajos, pero se alcanza la temperatura y humedad relativa planteada por la regla de asociación.

Figura 8
Árbol de decisión aplicado en la tercera regla de asociación



Conclusiones

A lo largo de la investigación se aplicaron técnicas de minería de datos, siendo los árboles de decisión los que permitieron realizar un análisis del comportamiento de las variables meteorológicas y contaminantes atmosféricos, conjuntamente con las reglas de asociación de forma mensualizada. Así, los datos analizados fueron los recolectados por las estaciones meteorológicas en la ciudad de Cuenca, Ecuador, del año 2018. De esa manera, una de las etapas primordiales que se realizaron en esta investigación fue la discretización de los datos, debido a que fue necesario realizar una correcta modulación de los datos obtenidos para aplicar el algoritmo y obtener resultados eficientes.

Para concluir esta investigación es importante tener en cuenta que los árboles de decisión son algoritmos de aprendizaje supervisado no paramétricos que permiten predecir el valor de una variable objetivo mediante el uso de variables predictivas. Por lo tanto, al utilizar este método en esta investigación hizo posible un control periódico de la calidad de las asociaciones de contaminantes y variables meteorológicas por estacionalidad, siendo así la comprobación de las reglas de asociación en la mayoría de los meses estudiados cumpliéndose más de 70% de ellas.

Como un trabajo futuro, al tener un conjunto de datos atmosféricos más amplio se puede llegar a predecir mediante la aplicación de árboles de decisión el comportamiento de los contaminantes atmosféricos y presentar un tratamiento efectivo y preventivo ante factores de riesgo que puedan afectar negativamente la salud de la población.

Reconocimientos

Los autores desean agradecer al Vicerrectorado de Investigaciones de la Universidad del Azuay por el apoyo financiero y académico, así como a todo el personal de la escuela de Ingeniería de Ciencias de la Computación, y el Laboratorio de Investigación y Desarrollo en Informática (LIDI).

Referencias

- Ahmed, H. W., & Alamire, D. J. H. (2020). Labelling Data for Correlation Pollution Dataset by Using Machine Learning. *Journal of Physics: Conference Series*, 1530(1), 012028. <https://doi.org/10.1088/1742-6596/1530/1/012028>
- Alsagheer, R. H. A., Alharan, A. F. H., & Al-Haboobi, A. S. A. (2017). Popular Decision Tree Algorithms of Data Mining Techniques: A Review. *International Journal of Computer Science and Mobile Computing*, 6(6), 133-142.
- Althuwaynee, O. F., Balogun, A., & Al Madhoun, W. (2020). Air pollution hazard assessment using decision tree algorithms and bivariate probability cluster polar function: evaluating inter-correlation clusters of PM10 and other air pollutants. *GIScience & Remote Sensing*, 57(2), 207-226. <https://doi.org/10.1080/15481603.2020.1712064>
- Athira, V., Geetha, P., Vinayakumar, R., & Soman, K. P. (2018). DeepAirNet: Applying Recurrent Networks for Air Quality Prediction. *Procedia Computer Science*, 132, 1394-1403. <https://doi.org/10.1016/j.procs.2018.05.068>
- Birant, D. (2011). Comparison of Decision Tree Algorithms for Predicting Potential Air Pollutant Emissions with Data Mining Models. *Journal of Environmental Informatics*, 17(1), 46-53. <https://doi.org/10.3808/jei.201100186>
- García-Ojeda, J. C., Alvites, J. A., & Puello, P. (2020). Cuantificación de la contaminación por dióxido de carbono producida por empresas courier usando minería de datos: una mirada a las provincias de Lima y Constitucional del Callao (Colombia). *Revista ESPACIOS*, 41(41), 1015.
- Gayathri, M., Shankar, R., & Duraisamy, S. (2020). Air pollution prediction using data mining technique. *International Research Journal of Engineering and Technology (IRJET)*, 7(8), 4292-4297.
- Hemada, B., & Vijaya Lakshmi, K. S. (2013). A Study On Discretization Techniques. *International Journal of Engineering Research & Technology*, 2(8).
- Huang, C.-J., & Kuo, P.-H. (2018). A Deep CNN-LSTM Model for Particulate Matter (PM2.5) Forecasting in Smart Cities. *Sensors*, 18(7), 2220. <https://doi.org/10.3390/s18072220>
- Manisalidis, I., Stavropoulou, E., Stavropoulos, A., & Bezirtzoglou, E. (2020). Environmental and health impacts of air pollution: a review. *Frontiers in public health*, 8, 14.
- Martínez-España, R., Bueno-Crespo, A., Soto, J., Muñoz, A., & Cecilia, J. (2018). Air-Pollution Prediction in Smart Cities through Machine Learning Methods: A Case of Study in Murcia, Spain. *Journal of Universal Computer Science*, 24(3), 261-276.

- Orellana, M., Salto, J., & Cedillo, P. (2021). Behavior Analysis of Atmospheric Components and Meteorological Variables Applying Data Mining Association Techniques. En *Proceedings of the 2021 Future of Information and Communication Conference (FICC)* (Vol. 2, pp. 192-204). Springer, Cham. https://doi.org/10.1007/978-3-030-73103-8_12
- Siwek, K., & Osowski, S. (2016). Data mining methods for prediction of air pollution. *International Journal of Applied Mathematics and Computer Science*, 26(2), 467-478. <https://doi.org/10.1515/amcs-2016-0033>
- Whiteman, C. D., Hoch, S. W., Horel, J. D., & Charland, A. (2014). Relationship between particulate air pollution and meteorological variables in Utah's Salt Lake Valley. *Atmospheric Environment*, 94, 742-753. <https://doi.org/10.1016/j.atmosenv.2014.06.012>

Metodología para localizar la Zona de Convergencia Intertropical usando velocidad de viento

Methodology to locate the Intertropical Convergence Zone using wind velocity

Jorge Bravo¹ <https://orcid.org/0009-0006-0466-9328>, Jesus Portilla¹ <https://orcid.org/0000-0003-4985-4325>

¹Escuela Politécnica Nacional, Quito, Ecuador

jdbravome@gmail.com, jesus.portilla@epn.edu.ec



Esta obra está bajo una licencia internacional
Creative Commons Atribución-NoComercial 4.0

Enviado: 2023/07/15

Aceptado: 2023/08/22

Publicado: 2023/10/15

Resumen

La localización de la Zona de Convergencia Intertropical (ITCZ) se puede establecer en función de la convergencia de corrientes de vientos alisios provenientes del hemisferio norte y sur. Esta zona tiene una gran relevancia climática pues incide directamente en la distribución de precipitaciones en las regiones ecuatoriales y tropicales a escala global. En este trabajo se desarrolla una metodología para localizar la ITCZ, empleando un algoritmo de convergencia en el campo vectorial de velocidades de viento obtenido del modelo de reanálisis atmosférico ERA Interim. Esta metodología resulta única, pues parte de un concepto robusto de análisis de gradientes y consigue un rango de la localización de la ITCZ en períodos de tiempo registrados históricamente, lo cual no se ha encontrado en estudios previos. La metodología consiste en una serie de algoritmos individuales que permiten limpiar, segmentar y depurar los datos de convergencia hasta obtener mapas que simplifican su visualización empleando MATLAB®. Se obtienen como resultado rangos estadísticos de la localización de la ITCZ en cada mes de 41 años de datos disponibles, patrones de desplazamiento meridional, y también mapas de contraste entre la localización mensual acumulada y la localización en meses de años en los que se manifestó El Niño Oscilación del Sur (ENSO) para mostrar una posible correlación.

Palabras clave: vientos alisios, ENSO, ITCZ, dinámica atmosférica, reanálisis.

Sumario: Introducción, Datos y Métodos, Resultados, Discusión de Resultados y Conclusiones.

Como citar: Bravo, J. Portilla, J. (2023). Metodología para localizar la Zona de Convergencia Intertropical usando velocidad de viento. *Revista Tecnológica - Espol*, 35(2), 61-75. Recuperado a partir de <http://www.rte.espol.edu.ec/index.php/tecnologica/article/view/1054>

Abstract

The location of the Intertropical Convergence Zone (ITCZ) can be established as a function of convergence between tropical trade winds from the Northern and Southern Hemispheres. This zone has a great impact on the weather since it directly affects the distribution of precipitation in equatorial and tropical regions on a global scale. This document elaborates on the development of a methodology to locate the ITCZ using a convergence algorithm in the vector field of wind velocity obtained from the ERA Interim reanalysis model. This methodology is unique because it relies on a robust concept of gradient analysis and achieves a location range for ITCZ in defined periods of time registered historically, which has not been developed in previous studies. The methodology consists of a series of individual algorithms that allow for the cleaning, segmentation, and debugging of convergence data, until simplified maps for visualization are obtained using MATLAB®. The results are statistical ranges of the location of the ITCZ in each month of the 41 years of data available, meridional shift patterns, and contrast maps between cumulative monthly location and the location on months of years where El Niño Southern Oscillation (ENSO) was present to show possible correlation.

Keywords: trade winds, ENSO, ITCZ, atmospheric dynamics, reanalysis.

Introducción

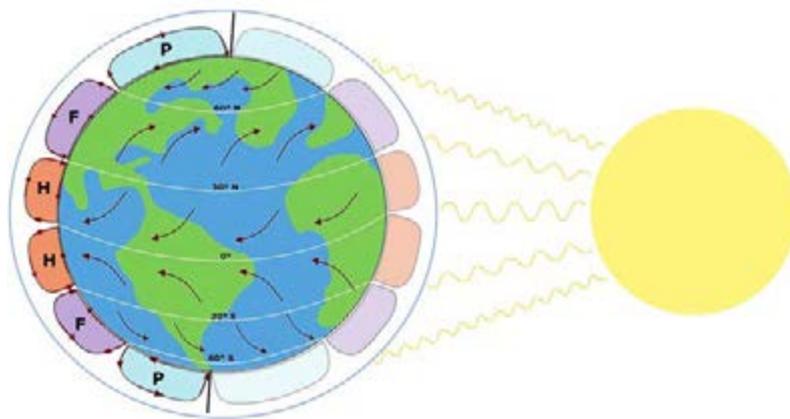
Muchas características del clima tropical son manifestaciones de la convergencia dinámica y termodinámica de los hemisferios norte y sur. Esto incluye el afloramiento ecuatorial (proceso por el cual agua fría profunda y rica en nutrientes asciende a la superficie), convección troposférica profunda (fenómeno térmico que permite el movimiento y mezcla de paquetes de aire a diferentes alturas de la atmósfera), una estructura compleja de corrientes oceánicas y uno de los esquemas de viento más estables del planeta, como lo son los vientos alisios (Waliser & Gautier, 1993). De todos los fenómenos que caracterizan a la región tropical, la característica que representa de forma más clara la intersección de los dos hemisferios es la Zona de Convergencia Intertropical, ITCZ por sus siglas en inglés, cuyo análisis es fundamental para entender la interacción entre la temperatura superficial del océano, precipitación, y presión a nivel del mar. La característica principal de la ITCZ es el alto contenido de vapor de agua en el aire, el cual al llegar a una mayor altura se enfría y provoca intensas lluvias, liberando humedad acumulada (Wallace & Hobbs, 2006).

La diferencia en presión atmosférica, generada por la variación de temperatura al receptor radiación solar en las diferentes latitudes del planeta, da lugar al movimiento de corrientes de aire (Marshall & Plumb, 2008). Las corrientes de aire se conocen como ‘vientos’ y siempre se desplazan desde una zona de presión alta a una zona de presión baja, y mientras mayor sea la diferencia de presiones, más fuerte será la velocidad del viento. Diferentes regiones de la superficie terrestre reciben diferentes cantidades de radiación, esto debido a la diferencia en el ángulo de incidencia, pero también al movimiento rotacional y estacional del planeta, variadas reflectividades de las superficies propias de las regiones y diversos factores de contaminación. De esta manera, la temperatura en distintas latitudes también varía. La región polar se caracteriza por ser una zona de baja temperatura y alta presión, mientras que la región ecuatorial se muestra como zona de alta temperatura y baja presión (Barry, 2003).

La diferencia de presión atmosférica y temperaturas produce una serie de patrones globales de circulación, los cuales componen el modelo de circulación atmosférica global, en el cual se establecen los ciclos de circulación como celdas de convección atmosférica. Existen tres celdas de convección atmosférica en cada hemisferio: Hadley, Ferrel y Polar (Barry, 2003).

En la Figura 1 se puede observar un esquema simplificado de la disposición de estas celdas en el planeta y el esquema de flujo de viento en la troposfera. Este estudio se enfoca en la región de la celda de convección atmosférica Hadley, que se caracteriza porque el aire cálido de la zona ecuatorial se eleva, creando un cinturón de baja presión, y mientras el aire se eleva va perdiendo calor, lo que resulta en condensación y da origen a las nubes y lluvia que se distribuyen al norte y sur de la línea ecuatorial. En las latitudes 30°N y 30°S, el aire frío y seco desciende, creando una región de alta presión que se caracteriza por ser despejada, y cuando alcanza la superficie de la tierra, se mueve hacia el trópico o hacia los polos (Barry, 2003). Por este motivo, en la zona ecuatorial hay largas cadenas de viento que transportan calor y humedad y dan lugar a la ITCZ, que es una franja compuesta por nubes de convección profunda ubicadas en una gran cadena longitudinal (Schneider et al., 2014).

Figura 1
Esquema de las celdas de convección atmosférica



Encontrar la localización de la ITCZ y sus propiedades migratorias usando datos de viento ha representado un desafío en el pasado por varias razones. Primero, las observaciones directas de la capa límite de vientos tropicales han sido escasas. Adicionalmente, las predicciones de rango corto, que usualmente proveen los datos para el análisis en predicción numérica del clima, han sido tradicionalmente menos exitosas en los trópicos que en las altitudes medias debido a la complejidad de la circulación tropical. Finalmente, las restricciones de balance aplicadas en la asimilación de datos presentaban poca información de viento del campo total de datos en los trópicos (Žagar et al., 2005). Para el estudio de la dinámica y climatología del flujo superficial en las áreas ecuatoriales, se han empleado datos de embarcaciones, sensores superficiales y sensores de altura; posteriormente se desarrollaron los reanálisis de datos y en años recientes se han tenido observaciones continuas de vectores de viento en la superficie oceánica gracias a dispersómetros basados en espacio, particularmente el escaterómetro (Žagar et al., 2011). Los análisis y pronósticos en los trópicos han mejorado significativamente en los últimos años, esto gracias a que se han desarrollado nuevas metodologías, más completas, que cuentan con la implementación de modelos variacionales para asimilación de datos, como los filtros de ensamble de Kalman, el 3D-Var y 4D-Var, los mismos que permiten mejorar las condiciones iniciales de pronóstico y su ajuste respecto al tiempo (Daescu, 2008; Houtekamer & Mitchell, 1998).

En este trabajo se propone una metodología que permita determinar la ubicación de la ITCZ mediante el análisis de los datos de velocidad del viento proporcionados por el Centro Europeo de Predicción del Clima de Rango Medio, ECMWF por sus siglas en inglés. Su archivo de reanálisis, ERA Interim, ha sido depurado de errores y presenta una mejor asimilación de

datos mediante el uso de nuevos parámetros para facilitar su procesamiento (Dee et al., 2011). Aportando con la metodología para ubicar la ITCZ se busca conseguir un contraste a estudios previos de su localización, bifurcaciones y zonas dobles de convergencia, y su variabilidad respecto al tiempo.

El uso acertado de la predicción meteorológica y datos climáticos permite una mejor planificación para varios sectores que aprovechan el océano para actividades económicas y permite establecer medidas preventivas contra fenómenos naturales como El Niño-Oscilación del Sur (ENSO) (Adams et al., 1995). En estudios anteriores se ha realizado un análisis histórico en el cual se observa que el desplazamiento hacia el norte de la ITCZ se ve acompañado de gradientes zonales de temperatura que responden a la presencia de ENSO (Koutavas et al., 2006), adicionalmente se ha observado que una de las zonas de mayor dificultad para la predicción de ENSO es la zona ecuatorial del Océano Pacífico este, una zona donde la ITCZ tiende a presentar bifurcaciones y variabilidad anómala (Mu et al., 2014; Sulca et al., 2018). Aunque es usual que el estudio de ENSO esté relacionado con una variación de la relación permanente que existe entre la intensidad de los vientos alisios y la temperatura superficial del océano zonal (McPhaden et al., 2006), es posible que la localización de la ITCZ presente anomalías que permitan identificar la presencia de condiciones de ENSO o brinden un parámetro adicional para mejorar la calidad predictiva de este fenómeno.

Datos y Métodos

Los datos empleados en este estudio permiten obtener información completa y verificada de las condiciones atmosféricas, oceanográficas y climáticas a escala global. Los modelos de reanálisis atmosférico consisten en una combinación de observaciones y medidas, con modelos físicos de predicción meteorológica. De esta manera, se obtiene como resultado un conjunto global de datos con mayor consistencia y menos vacíos en el tiempo. La mayor ventaja de este método es que emplea todas las variables posibles para construir un mapa global de datos, y con ello se consigue incrementar tanto la resolución como la calidad de las predicciones (Berrisford et al., 2011).

Para realizar un modelo de reanálisis se requiere establecer condiciones iniciales de los parámetros a evaluar y formular un modelo físico que tome en consideración el comportamiento de las variables atmosféricas en el tiempo. Las condiciones iniciales se obtienen de observaciones y mediciones de distintos instrumentos distribuidos en el Sistema Global de Observaciones, y se combinan con predicciones meteorológicas de corto rango. Esto permite ajustar los modelos predictivos luego de verificar sus resultados con el clima real (Berrisford et al., 2011). En este proceso se emplean métodos de asimilación de datos variacionales que permiten obtener una estructura de datos dinámica, que elimina errores de predicción y retroalimenta constantemente la base de datos inicial para mejorar la calidad de los resultados (ECMWF, 2017).

Actualmente se usan los datos de velocidad de viento a una altura de 10m, los mismos que se extraen de la base del ECMWF y tienen el respaldo de haber sido empleados en diversos proyectos para monitorear el cambio climático, planteamiento de decisiones estratégicas en industria y gobiernos, y propuestas de desarrollo de energía renovable (ECMWF, 2017). En este proyecto se usaron principalmente datos del modelo ERA Interim, que tienen una resolución espacial de 79 km, con intervalos temporales de 6 horas. Adicionalmente se plantea el uso futuro del modelo ERA 5, el cual tiene una resolución de 31km e intervalos temporales de 3 horas y además presenta mejoras significativas en calidad de los datos gracias a los avances en el método de reanálisis empleado por el ECMWF.

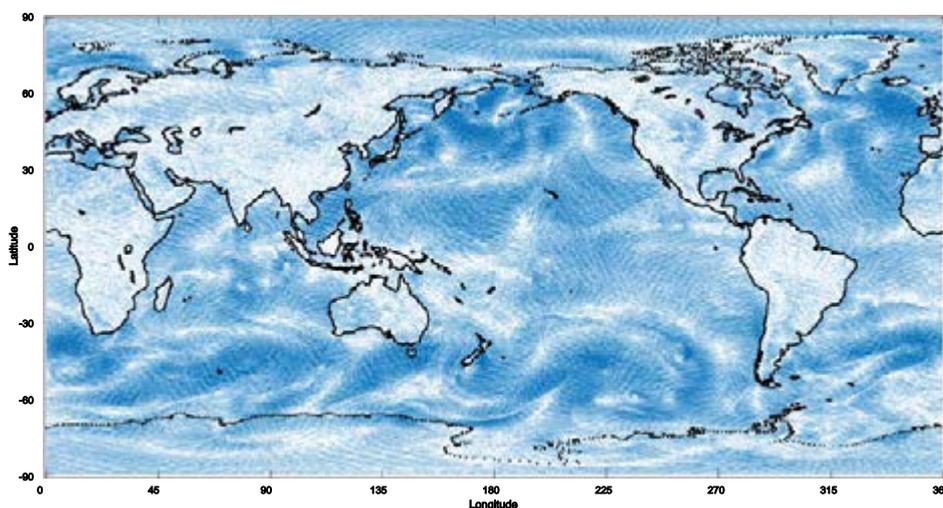
La metodología combina el análisis matemático de datos con herramientas de visualización computacional y estadística para estimar la localización de la ITCZ en diferentes rangos de tiempo. Para ello se crean algoritmos independientes, funciones y herramientas visuales, y se emplea como herramienta computacional el programa MATLAB®. La metodología consta de 3 pasos: el algoritmo de convergencia del viento, la agrupación y depuración de datos de convergencia, y el análisis estadístico de la localización de la ITCZ.

Diferentes proyectos meteorológicos, ambientales y energéticos han hecho uso de la ciencia de datos en combinación con modelos numéricos y datos satelitales para obtener resultados estadísticos (Gagne, 2016; García-Rojo, 2004).

Algoritmo de convergencia de viento

Para desarrollar el algoritmo de convergencia se consideran las características físicas que constituyen a la ITCZ, las cuales indican que es el resultado de la convergencia entre corrientes de viento alisios, opuestos en sentido meridional, debido a la diferencia de presión en diferentes latitudes. Esto permite aprovechar los datos de viento (velocidad a 10m de altitud), que se observan graficados en la Figura 2, y son provistos por los modelos de reanálisis para detectar esta convergencia en la inversión del sentido de flujo del viento. Esta inversión se identifica a través de un algoritmo de gradiente de imagen, el cual de forma simplificada, encuentra el cambio de sentido en cada celda de la malla donde se ubica el campo vectorial de velocidades (Mlsna & Rodriguez, 2009), y se aplica de manera bidireccional para evitar sesgos o desplazamientos de los resultados. Adicionalmente se emplea un criterio de ponderación, considerando que la velocidad del viento sea baja, sobre todo en su componente meridional, pues esta es otra característica de la ITCZ (Back & Bretherton, 2005).

Figura 2
Campo vectorial de velocidades de viento a 10m de altitud

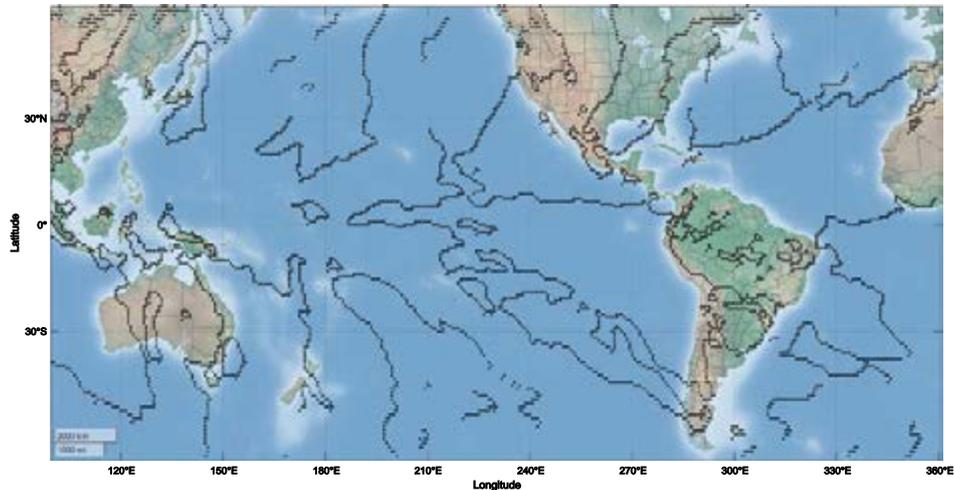


Por lo tanto, el algoritmo considera 3 condiciones para identificar puntos de convergencia:

- Componente meridional de la velocidad de viento menor a 10 m/s.
- Rapidez de viento menor a 10m/s.
- Cambio de dirección en la velocidad, representado por la inversión del sentido entre celdas adyacentes de la malla gaussiana.

El resultado del algoritmo es una matriz de puntos de convergencia, como se puede observar en la Figura 3. Estos puntos serán el punto de partida para la agrupación y depuración de datos.

Figura 3
Puntos de convergencia resultantes del algoritmo

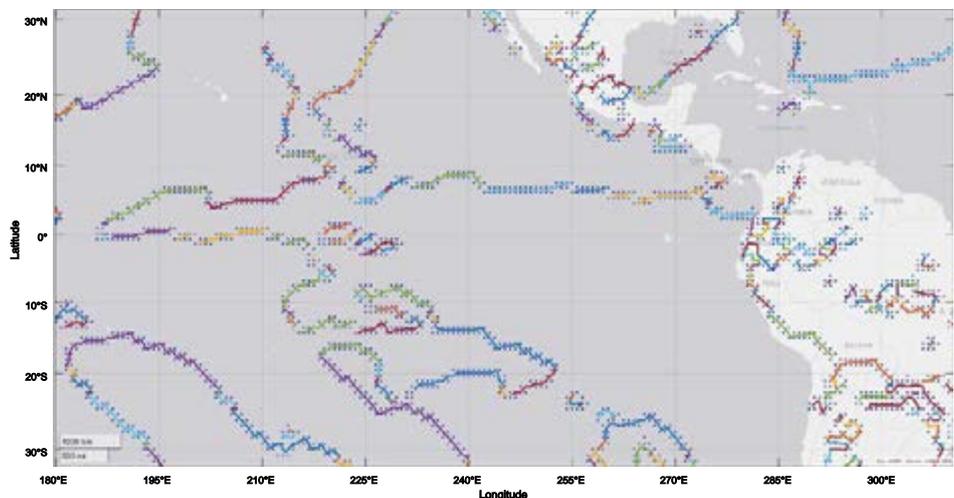


Agrupación y depuración de datos de convergencia

La agrupación de datos consiste en reducir la redundancia evitando perder la tendencia de los puntos de convergencia; es decir, se busca tener menos puntos que puedan mantener la misma información sobre la localización de la ITCZ.

Para este propósito se crea un algoritmo de agrupación que busca centroides para un grupo de puntos, que deben encontrarse dentro de una distancia máxima. La distancia máxima se calcula dentro del algoritmo en función de la distribución normal de la distancia entre dos puntos para todos los datos de convergencia. Se realizan varias iteraciones de esta agrupación, hasta que no existan puntos que se encuentren dentro de la distancia máxima, y que puedan agruparse. Dejando como resultado un esquema de puntos completamente equivalente al esquema inicial, pero con una cantidad reducida de datos, como puede observarse en la siguiente Figura 4.

Figura 4
Resultado de algoritmo de agrupación de puntos de convergencia

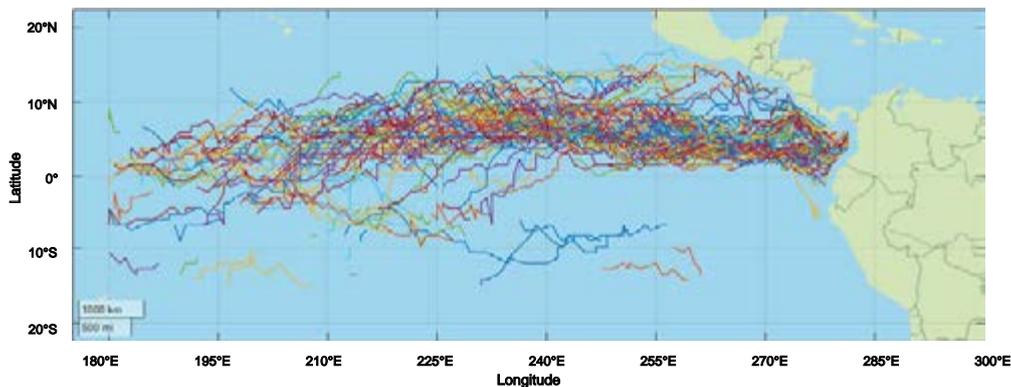


Posteriormente se realiza una segmentación de los datos, considerando una línea de base. Para identificar la línea de base se preseleccionan las líneas que se encuentran en el Pacífico oriental con mayor cercanía al continente americano. Esto debido a que existe una gran influencia de la región montañosa en los flujos de viento de la región (Xu et al., 2005). Adicionalmente, la bifurcación de la ITCZ con la Zona de Convergencia del Pacífico Sur (SPCZ) se manifiesta en un rango de longitudes más distante al continente (Bellucci et al., 2010)(Vincent, 1994). Partir desde una longitud próxima al continente facilita la identificación de la ITCZ en regiones de bifurcación.

Finalmente, luego de tener la línea de base como guía, es posible realizar una conexión de líneas adyacentes, siempre y cuando cumplan el criterio de unidireccionalidad y en caso de bifurcaciones, se una la línea que tenga mayor proximidad dentro de un rango predefinido bajo el criterio de distribución de distancia entre puntos. La línea obtenida como localización de la ITCZ para cada intervalo de tiempo, no es de gran utilidad por sí sola, pues únicamente contiene datos de un tiempo específico, por este motivo se debe agrupar de forma mensual, como se observa en la Figura 5, para poder obtener información relevante usando estadística.

Figura 5

Visualización de líneas de convergencia agrupadas para un mes

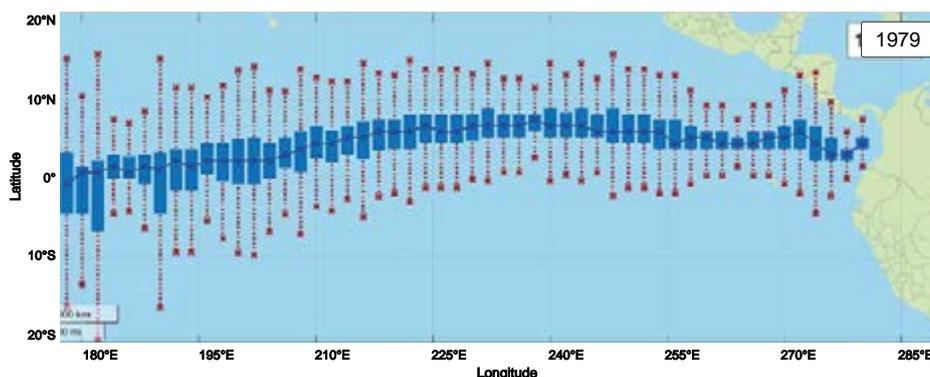


Análisis estadístico de localización de la ITCZ

Para realizar el análisis estadístico inicialmente se emplearon diagramas de cajas con los datos de convergencia acumulados de forma mensual. De esta forma se esperaba observar las zonas donde existe acumulación de puntos en un rango bien definido. Sin embargo, la amplitud de los rangos intercuartílicos y la dispersión de datos por valores atípicos hacen que este análisis no sea tan robusto, como se puede apreciar en la siguiente figura.

Figura 6

Diagrama de cajas para un mes de datos de convergencia acumulados

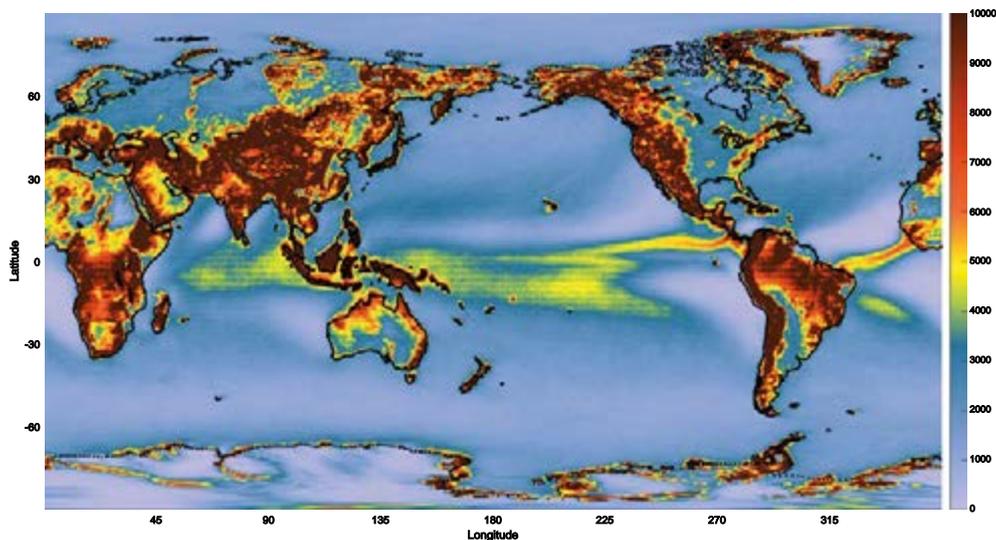


En esta sección se añade únicamente un diagrama de caja mensual como ejemplo. Sin embargo, se han obtenido resultados para los 12 meses, de los 41 años de datos disponibles en el modelo de reanálisis ERA Interim. Para facilitar su acceso digital, los gráficos se encuentran disponibles en la dirección: <https://doi.org/10.6084/m9.figshare.19102898.v1>.

Para contrastar el análisis estadístico de los diagramas de cajas, se aplica otro método, el cual consiste en obtener un conteo de datos en cada celda de la malla gaussiana para períodos acumulados de tiempo. De esta manera es posible visualizar dónde existe una gran acumulación de puntos y donde no hay convergencia, como se puede apreciar en la figura 7. Es también una visualización mucho más robusta de la convergencia en diferentes períodos de tiempo, pero a su vez no se consigue distinguir la ITCZ en zonas de bifurcación en el Océano Pacífico este. Por lo cual este paso aún se encuentra en desarrollo a fin de pulir los resultados finales.

Figura 7

Conteo de datos de convergencia acumulados para enero por 10 años



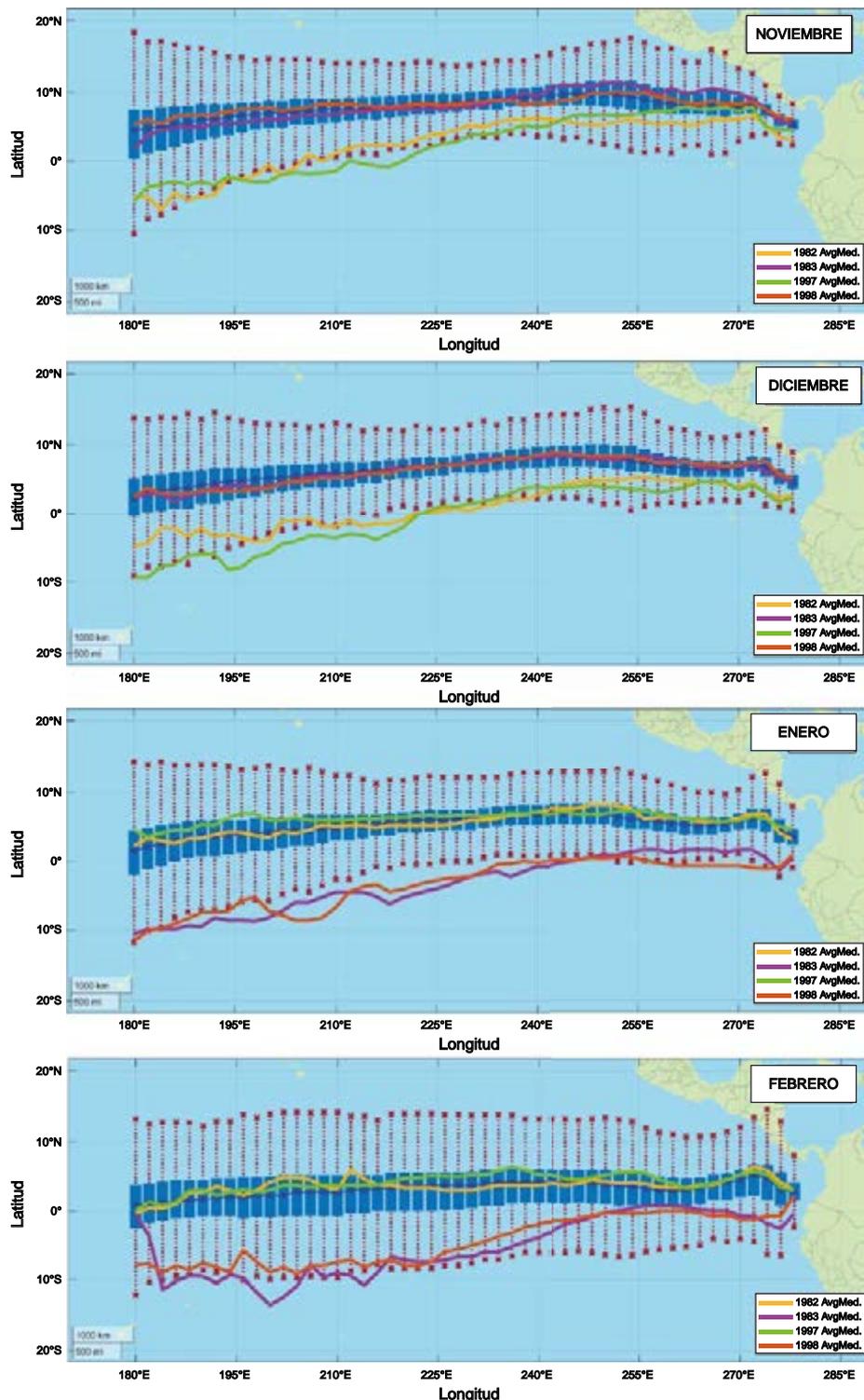
Resultados y Discusión

Mediante la distribución de los diagramas de cajas se visualiza, en cada mes, el rango de latitudes donde se encuentra la ITCZ, para cada intervalo de longitudes, y además las dimensiones del rango intercuartílico y el rango de límites externos resultan útiles como indicadores de la amplitud de la ITCZ. Como se observa en la figura 6, mientras mayor sea el rango intercuartílico (dimensión meridional de las cajas azules) y con ello el rango de límites externos, se interpreta que la amplitud de la ITCZ será mayor en esa región, pues la incertidumbre en su posición es mayor.

El uso de una metodología heurística en el desarrollo del estudio permite que los resultados tengan una aproximación distinta en comparación con métodos propuestos por diferentes autores. Aunque se ha planteado anteriormente la relación de la velocidad del viento con la formación de zonas de convergencia, sobre todo estableciendo su relación con la precipitación característica de las mismas, no existía un método de análisis computacional que empleara esta variable para localizar la zona de convergencia intertropical (Back & Bretherton, 2005). También es necesario considerar que los modelos de reanálisis brindan datos más precisos y compuestos en comparación con los anteriores esquemas de predicción meteorológica, por lo cual la metodología propuesta plantea criterios robustos para estimar la convergencia y relacionarla con la ubicación de la ITCZ, reduciendo la posibilidad de una acumulación de errores de cálculo o sesgos en los resultados (Bechtold et al., 2018).

Figura 8.

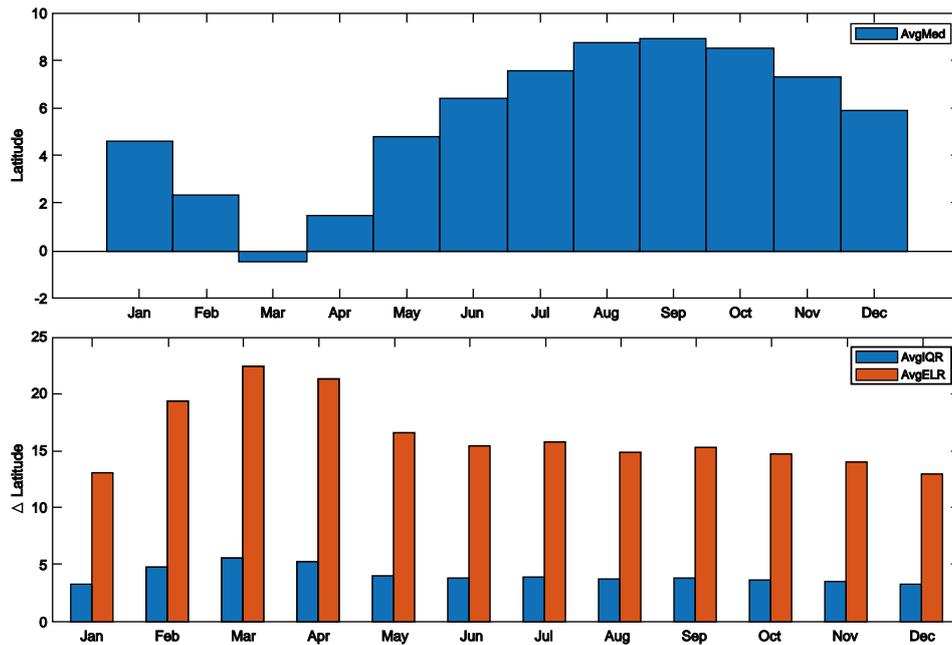
Diagramas de caja acumulados para un mismo mes en todos los años con datos disponibles en contraste con años con presencia de ENSO



Los diagramas de caja mensuales permiten observar que la distribución de posiciones de la ITCZ se mantiene mayoritariamente en un rango de latitudes entre 20°N y 15°S, y también se nota un desplazamiento hacia el norte de la línea ecuatorial acompañado con una disminución en el rango intercuartílico que se manifiesta de manera paulatina a partir del mes de abril.

Figura 9

Distribución de mediana, rango intercuartílico y de valores extremos.



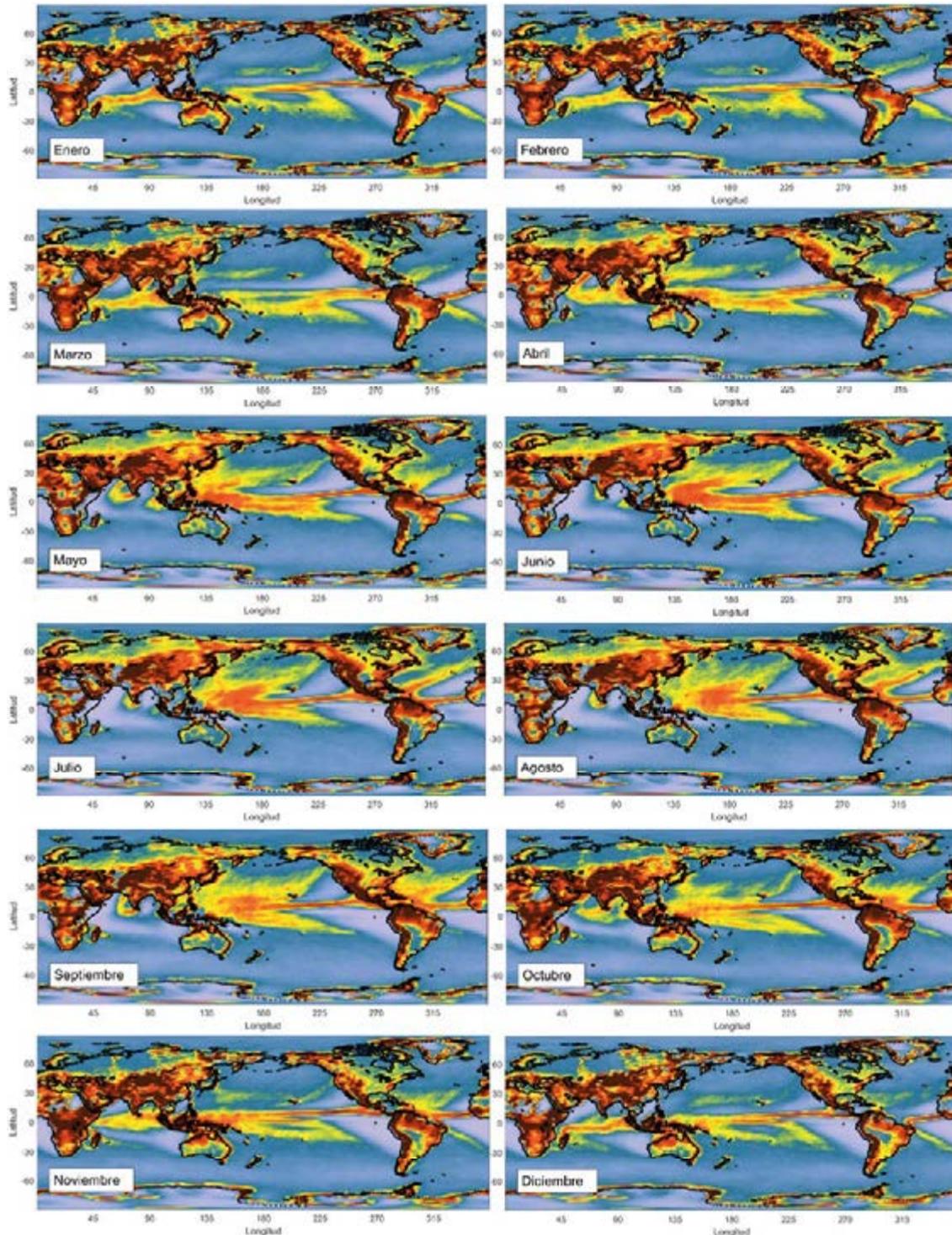
Esto se observa de forma resumida en la Figura 9, en la cual el uso de valores promedio para la distribución de cuartiles y los límites externos permite identificar una tendencia marcada del desplazamiento meridional de la ITCZ, alcanzando su valor más bajo en el mes de marzo, y su valor más alto entre agosto y septiembre. De esta forma, es posible coincidir con lo expuesto por Schneider (2014) y Philander (1996) en que la ITCZ se ubica principalmente sobre la línea ecuatorial, con puntuales excepciones en los meses de marzo y abril de ciertos años. A su vez, la distribución obtenida para la ubicación de la ITCZ en cada mes compone un recurso importante para su análisis, abasteciendo de datos y reduciendo la incertidumbre con otras zonas de convergencia o la existencia de bifurcaciones que mencionan Bellucci (2010) y Bischoff (2016).

Aunque el comportamiento de la ITCZ es consistente para cada mes a través de los años, existen variaciones abruptas importantes en los años donde se encontró activo el ENSO en la región. En la Figura 8 es posible visualizar un comportamiento atípico, en referencia a la distribución acumulada, entre los meses finales de 1982 hasta los primeros de 1983 y de manera similar entre los años de 1997 y 1998. Esto concuerda con lo planteado por Münnich (2005) y Ham (2014) respecto al descenso latitudinal de la ITCZ en relación con el flujo energético en la zona intertropical y su relación con la presencia del ENSO. También puede atribuirse el desplazamiento abrupto de la ITCZ a distintos factores medioambientales y de cambio climático como mencionan Hu (2007) y Sulca (2018), lo cual podría explicar la presencia de varios meses atípicos identificados en las distribuciones mensuales, como son: Enero 1992 y 2016, Febrero 1992, 2000, 2010 y 2017, Abril 1996 y 2012, entre otros.

Adicionalmente, en la Figura 10 se puede observar el desplazamiento de la ITCZ en diferentes meses del año, estos gráficos muestran la acumulación de puntos de convergencia para cada mes, en los 41 años de datos disponibles en el modelo de reanálisis ERA Interim. Se evidencia que existen desplazamientos más notorios en la región del Pacífico oriental que más dista del continente americano y desplazamientos cortos y menos frecuentes, en la región más cercana. También es posible identificar una variación estacional en esta figura, debido a que en cada año existe un comportamiento similar respecto a la ubicación en los diferentes

meses, cumpliendo con la tendencia a descender en el primer trimestre del año y luego ascender con fluctuaciones en el segundo y tercer trimestre, lo que ya había sido propuesto en estudios realizados por Lashkari (2017) y Keshtgar (2020).

Figura 10
Conteo de datos de convergencia acumulados para registro histórico



Al visualizar las distribuciones mensuales de todos los años se identifican meses con un rango intercuartílico atípico como: Febrero de 1984, 1995, Septiembre de 1982, 1992 y 2015, y Julio de 1994, entre otros. También es posible identificar una reducción del rango intercuartílico en los años posteriores a 1900, con excepciones puntuales, lo cual coincide con las proyecciones

de Byrne (2018) respecto a la disminución de la amplitud de la ITCZ en el siglo 20. La presencia puntual de valores atípicos en la variación del rango intercuartílico en los años con presencia del ENSO podría implicar que junto con el descenso longitudinal también se manifieste un incremento de la amplitud en la ITCZ como consecuencia de este fenómeno.

Sobre los resultados obtenidos también se puede interpretar que existe un patrón muy marcado y constante en la localización de la ITCZ. Como se observa en la Figura 10, es posible identificar varias zonas de convergencia importantes en el mapa global, bien definidas y delimitadas entre sí, aunque la excepción está en la zona del pacífico oriental, donde varios autores han identificado una zona de bifurcación por la gran masa de agua del océano pacífico que incide en la variabilidad de condiciones climáticas de la zona (Bischoff & Schneider, 2016; Keshtgar et al., 2020). Es posible afirmar que, en el registro histórico, los meses finales del año presentan una menor dispersión de datos, lo cual simplifica la localización de la ITCZ y reduce la incertidumbre; mientras que en contraste, los meses intermedios tienen un mayor valor para su rango intercuartílico promedio, dispersión y valores extremos por lo que su localización representa una dificultad mayor.

Finalmente, es necesario contextualizar que los resultados de esta metodología buscan ser un complemento para una serie de herramientas climatológicas de predicción, desarrolladas con la finalidad de integrar un sistema de soporte técnico para tomar decisiones respecto a la planificación y desarrollo de operaciones marinas en el país. Por esto, no se pueden considerar los resultados como un punto final, sino al contrario, como un punto de partida para continuar desarrollando un modelo completo que permita emplear la ITCZ como característica predictiva de clima y fenómenos atmosféricos y oceánicos. A su vez en el desarrollo de este estudio se encontraron avances respecto a la localización de la ITCZ en diferentes regiones, los mismos que emplean modelos de datos atmosféricos similares combinados con un procesamiento estadístico, como el trabajo de Žagar (2011), Colna (2017), y Lashkari (2017).

Conclusiones

La elaboración de un algoritmo para el post-procesamiento, filtrado y visualización de datos de viento es posible desde diferentes enfoques; en este estudio el uso de un método heurístico simplificado permitió que el tiempo de trabajo del algoritmo y la memoria requerida para almacenar los datos fueran cortos. Adicionalmente, el uso de herramientas gráficas y estadísticas combinadas representó una gran ventaja para analizar todas las variables necesarias e incrementar la confiabilidad de los resultados.

En los algoritmos de procesamiento se realizaron varias iteraciones a fin de modelar e identificar los rangos en los que se ubican diferentes zonas de convergencia. Este planteamiento, junto con herramientas visuales de registro acumulado de tiempo, facilitó la identificación de las variables físicas y geográficas que permitieron establecer la localización de la ITCZ, especialmente en zonas como el Océano Pacífico Oriental donde la presencia de bifurcación entre zonas de convergencia o la doble ITCZ complicaba el análisis.

El criterio inicial para encontrar estadísticamente la localización de la ITCZ no es suficientemente robusto, considerando la presencia de valores atípicos y la dificultad de reducir la dispersión de datos al emplear únicamente algoritmos de conexión basados en rangos de distancia y caracterización de líneas de convergencia, por lo cual se propuso un contraste estadístico que facilite la visualización acumulada de datos. Esto ha abierto una nueva perspectiva al estudio, pues se podrán combinar ambos criterios para permitir delimitar la ITCZ en zonas de bifurcación y respaldar el análisis con el registro histórico acumulado, que muestra de forma clara la acumulación de puntos de convergencia con patrones constantes.

Aunque la resolución y confiabilidad de los datos del modelo de reanálisis ERA Interim es buena, sería mejor emplear todos los algoritmos en los datos mejorados del modelo de reanálisis ERA 5, pues esto permitirá incrementar la resolución de los resultados y a su vez obtener mayor precisión en las zonas que pueden ser de gran interés a una escala nacional o regional.

Reconocimientos

Este trabajo ha sido financiado parcialmente por el Proyecto de Investigación CEDIA IDI XVII-2023-12-WEC GALAPAGOS y EPN PIGR-2103. Los autores desean expresar su agradecimiento a CEDIA y la Escuela Politécnica Nacional.

Referencias

- Adams, R. M., Bryant, K. J., Mccarl, B. A., Legler, D. M., O'Brien, J., Solow, A., & Weiher, R. (1995). Value of Improved Long-Range Weather Information. *Contemporary Economic Policy*, 13(3), 10–19. <https://doi.org/10.1111/j.1465-7287.1995.tb00720.x>
- Back, L. E., & Bretherton, C. S. (2005). The relationship between wind speed and precipitation in the Pacific ITCZ. *Journal of Climate*, 18(20), 4317–4328. <https://doi.org/10.1175/JCLI3519.1>
- Barry, R. G. (2003). Atmosphere, Weather and Climate. In Atmosphere, *Weather and Climate*. Routledge. <https://doi.org/10.4324/9780203871027>
- Bechtold, P., Köhler, M., Jung, T., Doblas-Reyes, F., Leutbecher, M., Rodwell, M. J., Vitart, F., & Balsamo, G. (2018). Advances in simulating atmospheric variability with the ECMWF model: From synoptic to decadal time-scales. *Quarterly Journal of the Royal Meteorological Society*, 1227(July), 496. <https://doi.org/10.1002/qj>
- Bellucci, A., Gualdi, S., & Navarra, A. (2010). The double-ITCZ syndrome in coupled general circulation models: The role of large-scale vertical circulation regimes. *Journal of Climate*, 23(5), 1127–1145. <https://doi.org/10.1175/2009JCLI3002.1>
- Berrisford, P., Dee, D. P., Poli, P., Brugge, R., Fielding, M., Fuentes, M., Kållberg, P. W., Kobayashi, S., Uppala, S., & Simmons, A. (2011). The ERA-Interim archive Version 2.0. In *ERA Report Series (ERA Report)*. ECMWF. <https://www.ecmwf.int/node/8174>
- Bischoff, T., & Schneider, T. (2016). The equatorial energy balance, ITCZ position, and double-ITCZ bifurcations. *Journal of Climate*, 29(8), 2997–3013. <https://doi.org/10.1175/JCLI-D-15-0328.1>
- Byrne, M. P., Pendergrass, A. G., Rapp, A. D., & Wodzicki, K. R. (2018). Response of the Intertropical Convergence Zone to Climate Change: Location, Width, and Strength. *Current Climate Change Reports*, 4(4), 355–370. <https://doi.org/10.1007/s40641-018-0110-5>
- Colna, K. E. (2017). *Latitudinal position and trends of the intertropical convergence zone (ITCZ) and its relationship with upwelling in the southern Caribbean Sea and global climate indices*. University of South Florida.
- Daescu, D. N. (2008). On the Sensitivity Equations of Four-Dimensional Variational (4D-Var) Data Assimilation. *Monthly Weather Review*, 136(8), 3050–3065. <https://doi.org/10.1175/2007MWR2382.1>
- Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M. A., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A. C. M., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A. J., ... Vitart, F. (2011). The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. *Quarterly Journal of the Royal Meteorological Society*, 137(656), 553–597. <https://doi.org/10.1002/qj.828>

- ECMWF. (2017). *20 years of 4D-Var: better forecasts through a better use of observations*.
<https://www.ecmwf.int/en/about/media-centre/news/2017/20-years-4d-var-better-forecasts-through-better-use-observations>
- Gagne, D. J. I. I. (2016). *Coupling data science techniques and numerical weather prediction models for high-impact weather prediction*.
- García-Rojo, R. (2004). Algorithm for the estimation of the long-term wind climate at a meteorological mast using a joint probabilistic approach. *Wind Engineering*, 28(2), 213–224. <https://doi.org/10.1260/0309524041211378>
- Ham, Y.-G., & Kug, J.-S. (2014). Effects of Pacific Intertropical Convergence Zone precipitation bias on ENSO phase transition. *Environmental Research Letters*, 9(6), 64008. <https://doi.org/10.1088/1748-9326/9/6/064008>
- Houtekamer, P. L., & Mitchell, H. L. (1998). Data assimilation using an ensemble Kalman filter technique. *Monthly Weather Review*, 126(3), 796–811.
- Hu, Y., Li, D., & Liu, J. (2007). Abrupt seasonal variation of the ITCZ and the Hadley circulation. *Geophysical Research Letters*, 34(18). <https://doi.org/https://doi.org/10.1029/2007GL030950>
- Keshtgar, B., Alizadeh-Choobari, O., & Irannejad, P. (2020). Seasonal and interannual variations of the intertropical convergence zone over the Indian Ocean based on an energetic perspective. *Climate Dynamics*, 54(7–8), 3627–3639. <https://doi.org/10.1007/s00382-020-05195-5>
- Koutavas, A., deMenocal, P. B., Olive, G. C., & Lynch-Stieglitz, J. (2006). Mid-Holocene El Niño–Southern Oscillation (ENSO) attenuation revealed by individual foraminifera in eastern tropical Pacific sediments. *Geology*, 34(12), 993–996. <https://doi.org/10.1130/G22810A.1>
- Lashkari, H., Mohammadi, Z., & Keikhosravi, G. (2017). Annual Fluctuations and Displacements of Inter Tropical Convergence Zone (ITCZ) within the Range of Atlantic Ocean-India. *Open Journal of Ecology*, 07(01), 12–33. <https://doi.org/10.4236/oje.2017.71002>
- Marshall, J., & Plumb, R. A. (2008). Atmosphere, Ocean, and Climate Dynamics. In *Zhurnal Eksperimental'noi i Teoreticheskoi Fiziki*. <http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:No+Title#0>
- McPhaden, M. J., Zebiak, S. E., & Glantz, M. H. (2006). ENSO as an Integrating Concept in Earth Science. *Science*, 314(5806), 1740–1745. <https://doi.org/10.1126/science.1132588>
- Mlsna, P. A., & Rodriguez, J. J. (2009). Gradient and Laplacian edge detection. In *The essential guide to image processing* (pp. 495–524). Elsevier.
- Mu, M., Yu, Y., Xu, H., & Gong, T. (2014). Similarities between optimal precursors for ENSO events and optimally growing initial errors in El Niño predictions. *Theoretical and Applied Climatology*, 115(3), 461–469. <https://doi.org/10.1007/s00704-013-0909-x>
- Münnich, M., & Neelin, J. D. (2005). Seasonal influence of ENSO on the Atlantic ITCZ and equatorial South America. *Geophysical Research Letters*, 32(21), 1–4. <https://doi.org/10.1029/2005GL023900>
- Philander, S. G. H., Gu, D., Halpern, D., Lambert, G., Lau, N. C., Li, T., & Pacanowski, R. C. (1996). Why the ITCZ is mostly north of the equator. In *Journal of Climate* (Vol. 9, Issue 12, pp. 2958–2972). [https://doi.org/10.1175/1520-0442\(1996\)009<2958:WTIIMN>2.0.CO;2](https://doi.org/10.1175/1520-0442(1996)009<2958:WTIIMN>2.0.CO;2)
- Schneider, T., Bischoff, T., & Haug, G. H. (2014). Migrations and dynamics of the intertropical convergence zone. *Nature*, 513(7516), 45–53. <https://doi.org/10.1038/nature13636>
- Sulca, J., Takahashi, K., Espinoza, J.-C., Vuille, M., & Lavado-Casimiro, W. (2018). Impacts of different ENSO flavors and tropical Pacific convection variability (ITCZ, SPCZ) on austral summer rainfall in South America, with a focus on Peru. *International Journal of Climatology*, 38(1), 420–435. <https://doi.org/https://doi.org/10.1002/joc.5185>

- Vincent, D. G. (1994). The South Pacific Convergence Zone (SPCZ): A Review. *Monthly Weather Review*, 122(9), 1949–1970. [https://doi.org/10.1175/1520-0493\(1994\)122<1949:TSPCZA>2.0.CO;2](https://doi.org/10.1175/1520-0493(1994)122<1949:TSPCZA>2.0.CO;2)
- Waliser, D. E., & Gautier, C. (1993). A satellite-derived climatology of the ITCZ. In *Journal of Climate* (Vol. 6, Issue 11, pp. 2162–2174). [https://doi.org/10.1175/1520-0442\(1993\)006<2162:ASDCOT>2.0.CO;2](https://doi.org/10.1175/1520-0442(1993)006<2162:ASDCOT>2.0.CO;2)
- Wallace, J. M., & Hobbs, P. V. (2006). Atmospheric Science An introductory survey. In *International Geophysics Series* (Second). Elsevier.
- Warf, B. (2014). Atmospheric Circulation. In *Encyclopedia of Geography*. Encyclopædia Britannica. <https://doi.org/10.4135/9781412939591.n63>
- Xu, H., Xie, S. P., Wang, Y., & Small, R. J. (2005). Effects of Central American Mountains on the Eastern Pacific Winter ITCZ and moisture transport. *Journal of Climate*, 18(18), 3856–3873. <https://doi.org/10.1175/JCLI3497.1>
- Žagar, N., Andersson, E., & Fisher, M. (2005). Balanced tropical data assimilation based on a study of equatorial waves in ECMWF short-range forecast errors. *Quarterly Journal of the Royal Meteorological Society*, 131(607), 987–1011. <https://doi.org/10.1256/qj.04.54>
- Žagar, N., Skok, G., & Tribbia, J. (2011). Climatology of the ITCZ derived from ERA Interim reanalyses. *Journal of Geophysical Research: Atmospheres*, 116(D15). <https://doi.org/10.1029/2011JD015695>

El modelo OSeMOSYS y la descarbonización a largo plazo del sector eléctrico en Galápagos

OSeMOSYS model and the long-term decarbonization of the power sector in Galapagos

Esteban Meneses^{1, 2} <https://orcid.org/0000-0002-8578-956X>, Rafael Soria² <https://orcid.org/0000-0003-4836-1510>

¹*Escuela Politécnica del Litoral, Quito, Ecuador*
estebanmeneses@gmail.com

²*Universidad San Francisco de Quito, Quito, Ecuador*
rsoriap@usfq.edu.ec



Esta obra está bajo una licencia internacional
Creative Commons Atribución-NoComercial 4.0

Enviado: 2023/07/16

Aceptado: 2023/08/22

Publicado: 2023/10/15

Resumen

OSeMOSYS (*Open Source Modelling System*) es un modelo de equilibrio parcial que busca satisfacer la demanda energética en el horizonte de tiempo planteado, que se optimiza por mínimo costo. El modelo ofrece diversas ventajas debido a su fácil uso y su condición de código abierto, lo que le permite estar en constante actualización y evolución gracias a una comunidad creciente. Entre sus aplicaciones están los sistemas energéticos, los diferentes usos del suelo y la capacidad de plantear distintos escenarios de acuerdo con la información disponible. El trabajo presenta los principales componentes del modelo, sus diversas opciones de uso y casos de uso en algunas regiones del mundo, incluido el caso de expansión eléctrica al año 2050 en las Islas Galápagos. Como resultado, no solo es posible contar con electricidad 100% renovable en las Islas, también es un escenario menos costoso en relación a mantener el sector eléctrico como se encuentra en la actualidad. Dicha inversión sería menos de la mitad de lo que se gastaría si se sigue con las políticas eléctricas actuales. Para alcanzar la descarbonización de las Islas, se tiene que empezar a invertir en energías renovables que, con el mejoramiento de su tecnología, son cada vez menos costosas y más eficientes. Es importante recalcar que, para cualquier plan a largo plazo, se debe contar con los estudios ambientales necesarios, más aun tratándose de un lugar con tanta riqueza natural.

Sumario: Introducción, Metodología, Resultados y Discusión y Conclusiones.

Como citar: Meneses, E. & Soria, R. (2023). El modelo OSeMOSYS y la descarbonización a largo plazo del sector eléctrico en Galápagos. *Revista Tecnológica - Espol*, 35(2), 76-93. Recuperado a partir de <http://www.rte.espol.edu.ec/index.php/tecnologica/article/view/1061>

Palabras clave: código abierto, planificación energética, energía renovable, capacidad, potencia, gases de efecto invernadero (GEI).

Abstract

OSeMOSYS (Open Source Modeling System) is a partial equilibrium model that aims to meet energy demand over a specified time horizon, optimizing for minimum cost. The model offers a variety of advantages due to its ease of use and open-source status, which allows it to be constantly updated and evolved by a growing community. Among its applications are energy systems as well as different possible land uses. In addition, the user can establish a variety of scenarios according to the available information. In this way, the work presents its main components as well as its various usage options, in addition to case studies in some regions of the world, including the case of electrical expansion to 2050 in the Galapagos Islands. As a result, it is possible to have 100% renewable electricity in the islands, and it is a less costly scenario than maintaining the electricity sector as it is today. This investment would be less than half of what would be spent if current electricity policies were continued. To achieve decarbonization on the islands, it is necessary to start investing in renewable energies, which, with the improvement of their technology, are becoming increasingly less expensive and more efficient. It is important to note that any long-term plan must include the necessary environmental studies, especially in a place with so much natural wealth.

Keywords: open source, energy planning, renewable energy, capacity, power, greenhouse gases (GHG).

Introducción

El archipiélago de las Galápagos se encuentra alrededor de 1.000 km de las costas de Ecuador. Estas islas son jóvenes en términos geológicos, lo que las hace vulnerables a especies con mayor tiempo de evolución. Su historia de asentamientos humanos comenzó a mediados del milenio pasado, cuando eran visitadas por piratas y balleneros, y después fueron colonizadas por primera vez para realizar actividades agrícolas por exiliados y presos políticos de los primeros gobiernos de la naciente República del Ecuador, en los años 30 del siglo XIX. Las islas fueron declaradas Patrimonio Natural de la Humanidad y Reserva de la Biósfera por la UNESCO¹ en el año de 1978 y 1984 respectivamente, desde entonces las islas han crecido demográficamente, al igual que el turismo que llega cada año a las islas del Parque Nacional Galápagos, PNG (2023).

Actualmente, el 97% del área del archipiélago es parte del PNG y la Reserva Marina de Galápagos. El 3% o 230 km² del área de las cuatro islas pobladas está destinada para asentamientos humanos; de esta área, el 48% está en la isla Santa Cruz, sobre todo en las zonas altas designadas para la agricultura (84%) (WWF, 2003). Demográficamente, Santa Cruz es la más poblada, con alrededor de 18.000 habitantes, seguida de San Cristóbal (7.200), Isabela (2.400) y Floreana (160). Estas islas concentran sus poblaciones alrededor de los puertos, con pequeños asentamientos en zonas agrícolas altas (INEC, 2015). La economía de las Islas Galápagos pasó de ser agrícola a una basada en el turismo, siendo este rubro el 88% de sus ingresos en 2006 (Epler 2007). Sin embargo, De Miras (1995) estimó que solo el 7% de la riqueza se queda en las islas, ya que los principales operadores turísticos son extranjeros o del territorio continental ecuatoriano.

1 Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura.

El número de visitantes a las Islas Galápagos aumentó significativamente desde finales de la década del ochenta. Así, el número de turistas que visitan las islas superó el límite oficial recomendado de 12.000 habitantes. Para hacer frente a la demanda creciente de turistas, el límite se amplió a 25.000, pero incluso este nuevo límite fue superado. Esto plantea preguntas sobre cuánto tiempo más podrán las islas soportar el crecimiento vertiginoso del turismo. En el año 2022, el archipiélago recibió 267 mil turistas (MAATE, 2023). Asimismo, la población local también experimentó un crecimiento exponencial, pasando de 1.346 habitantes en 1950 a 25.244 en 2015 (INEC, 2015). Este crecimiento de turistas y población local se reflejó en la economía, que creció aproximadamente un 78% entre 1999 y 2005, o un 9,6 % anual, lo que llevó a Galápagos a ser una de las regiones de mayor crecimiento a nivel mundial en términos económicos (Taylor, Hardner, and Stewart, 2009). El valor agregado bruto (VAB) en las islas en el año 2022 fue alrededor de USD 260 millones² y su VAB per cápita fue alrededor de USD 9 mil (BCE 2023a).

Este crecimiento turístico ha impulsado la inmigración, lo que provoca un riesgo para la seguridad alimentaria del lugar (Walsh and Mena, 2016). Según Sampedro et al. (2020), las islas importaron el 75% de los productos agrícolas que consumieron en el año 2017. Estos volúmenes las hacen vulnerables a las especies invasoras. Como resultado, dos de las especies de plantas invasoras, la guayaba y la mora, se han apoderado de áreas de tierra que antes se cultivaban, incluso en áreas restringidas del parque (Snell et al., 2002). Tanto el crecimiento demográfico como el turístico han generado preocupación en términos energéticos, por lo que resulta necesario considerar la expansión energética del archipiélago.

Actualmente, la electricidad en las islas Galápagos proviene en su mayoría de motores de combustión interna (84%), eólica (12%) y solar (4%). Las islas importan grandes volúmenes de derivados de hidrocarburos para satisfacer las necesidades energéticas de la población local y el turismo (IIGE, 2018). En consecuencia, la enorme riqueza biológica de las islas resulta vulnerable a posibles derrames de combustible. Es así como, en 2008, el gobierno ecuatoriano lanzó la *Iniciativa Cero Combustibles Fósiles en Galápagos*, con la finalidad de eliminar el uso de combustibles fósiles y disminuir la emisión de gases de efecto invernadero al ambiente (PNUD, 2016).

Para la construcción de un modelo de expansión eléctrica a largo plazo, se tienen varias opciones de modelos y programas, en su mayoría modelos pagados, lo que dificulta su uso masivo, principalmente en economías en desarrollo. En consecuencia, el modelo de código abierto OSeMOSYS³ representa una opción de fácil acceso y aprendizaje. Este modelo de equilibrio parcial, que optimiza al menor costo, considera: eficiencia tecnológica, infraestructura actual energética, capacidad de almacenamiento actual, la vida operacional de las tecnologías, factores de capacidad de las plantas, perfiles de recursos disponibles (viento, sol, etc), pérdidas por distribución y/o transmisión, variables exógenas al modelo como el PIB, crecimiento poblacional, entre otras. Dicho modelo fue presentado por primera vez en el taller IEA en 2008, escrito en GNU MathProg y después desarrollado en GAMS⁴ y Python (Niet et al., 2021). Esta herramienta fue desarrollada para ayudar a los planificadores de energía, así como para la comunidad científica de energía (Howells et al., 2011).

El trabajo tiene como objetivo mostrar el funcionamiento del modelo OSeMOSYS y

2 Valores provisionales.

3 OSeMOSYS (Open Source Modelling System).

4 General Algebraic Modelling System.

algunos casos de uso. Por tanto, el trabajo se estructura de la siguiente manera: La sección de metodología describe el modelo, sus principales características y plataformas de uso, y datos del caso en las islas; la sección de resultados indica los principales hallazgos de los casos a nivel global, y el caso de uso en las islas Galápagos; finalmente se presentan algunas conclusiones.

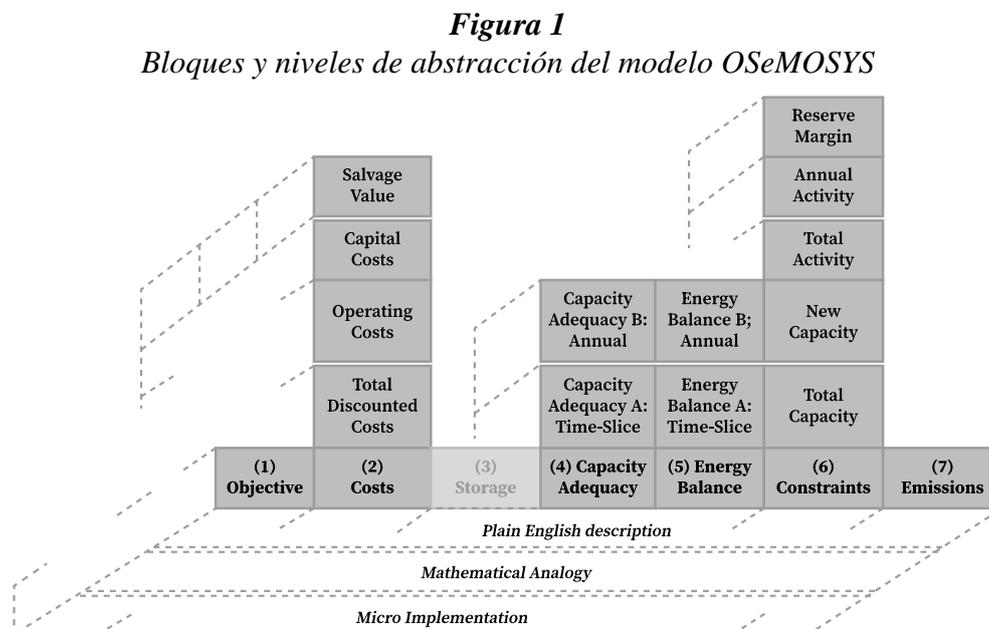
Metodología

Modelo OSeMOSYS

OSeMOSYS es un modelo de equilibrio, es decir, la oferta de energía es igual a la demanda. El modelo, al ser de código abierto, es un modelo que permite mejoras. Por tanto, es posible acceder a las ecuaciones del modelo y modificarlas a conveniencia, lo cual requiere un poco de destreza en la escritura de código.

Actualmente, existe también la nube de OSeMOSYS, donde el usuario puede crear una cuenta y correr el modelo, y asimismo se puede encontrar la última versión de la aplicación aquí. Tanto la aplicación como la nube de OSeMOSYS funcionan en conjunto; en la aplicación se ingresan los datos al modelo y en la nube se corre el modelo. El usuario tiene a disposición el programa *clicSAND* (Cannone et al. 2022), permite correr el modelo con un archivo Excel para mayor facilidad. Para más información, se pueden tomar los cursos en línea disponibles gratuitamente en Open University (University 2023).

El modelo está desarrollado en bloques y cada uno en tres niveles de abstracción. Los parámetros, las variables, las restricciones y las relaciones con sus objetivos están descritas en inglés, así como, la formulación algebraica y su implementación en un lenguaje de programación. Fue desarrollado en bloques para facilitar su independencia y mejoras. Existen siete bloques: (1) objetivo, (2) costos, (3) almacenamiento, (4) capacidad adecuada, (5) balance energético, (6) restricciones, (7) y emisiones (Howells et al., 2011). Ver Figura 1⁵.



Para la definición del modelo existe: Conjuntos (ej. rango temporal, intervalos de tiempo, etc.), parámetros globales, parámetros de demandas, parámetros de tecnología, parámetros de desempeño, parámetros de costos, parámetros de capacidad, restricciones de capacidad,

5 Figura 1, adaptada de (Howells et al., 2011).

restricciones de inversión, restricciones de actividad, márgenes de reserva, emisiones y penalidades. Para mayor detalle sobre los parámetros, variables, conjuntos y ecuaciones, remítase al manual OSeMOSYS (2023).

Función Objetivo

La función objetivo que se encuentra en el primer bloque (ver Figura 1) calcula el mínimo valor presente neto del sistema energético para satisfacer la demanda dada, que puede ser solo del sector eléctrico o uno que incluya todos los sistemas energéticos y usos de suelo. Para alcanzar la función objetiva deben satisfacer las diferentes restricciones (Howells et al., 2011).

Costos

Se tienen los costos de cada tecnología, tanto los de capital de inversión, así como los de operación y mantenimiento, fijos y variables. Además, se puede ingresar los provenientes de penalidades por emisiones de GEI. Los costos variables se calculan para cada intervalo de tiempo especificado por el usuario, modo de operación de la tecnología, año de operación y región modelada. Los costos fijos se calculan de acuerdo con la capacidad instalada de cada tecnología. Al final se obtiene el costo de operación y mantenimiento anual que es la suma de los costos fijos y variables, estos son llevados al primer año modelado con una tasa de descuento. Los costos anuales de inversión están determinados por el nivel de capacidad existente y nueva, e igualmente son llevados al primer año modelado con una tasa de descuento global o por tecnología (Howells et al., 2011).

Capacidad Adecuada

Debe existir una capacidad suficiente para cada tecnología que permita cubrir la demanda energética anual. Es necesario determinar la capacidad total disponible previa al periodo del modelo, es decir, la capacidad actual al comienzo del periodo. Para determinar la capacidad suficiente anual, se calcula la tasa de actividad de cada tecnología, intervalo de tiempo y región modelada que es ingresada por el usuario en los factores de capacidad (Howells et al., 2011).

Balance Energético

Los niveles de operación toman en cuenta la actividad, el uso de energía, la producción de energía y las emisiones en cada modo de operación y tecnología que son calculados por intervalos de tiempo. Por tanto, es importante que la producción energética cubra la demanda en cada intervalo. La tasa de producción de combustibles por tecnología es determinada multiplicando la tasa de actividad de la relación de entrada/salida del combustible, que es la eficiencia de la tecnología (Howells et al., 2011).

Restricciones

Podemos limitar el modelo de varias formas, mediante capacidad con límites superiores y/o inferiores, por generación con límites superiores y/o inferiores, por cantidad de emisiones por el periodo total del modelo y mediante penalidades por emisiones por tecnología y/o por región. Finalmente, se lo puede limitar por máxima o mínima inversión en términos monetarios por unidad de capacidad. Dichos límites se los puede hacer por tecnología, por región y por año (Howells et al., 2011).

Emisiones

Para las emisiones de GEI el usuario ingresa las relaciones de emisiones por unidad de actividad. Se puede considerar más de un GEI por cada tecnología y modo de operación; solo la actividad promedio de cada tecnología y su modo de operación se multiplican por la relación emisión/actividad. Las emisiones anuales se obtienen de la suma de las emisiones anuales por

modo de operación. Sin embargo, no todas las tecnologías son consideradas en las emisiones, por ejemplo, las tecnologías renovables como la solar y la eólica. (Howells et al., 2011).

OSeMOSYS Galápagos

Área de Estudio

El archipiélago de Galápagos está formado por 22 islas, entre grandes y medianas, y 107 islotes de origen volcánico con una inmensa riqueza biológica (WWF, 2003). Estas islas enfrentan un permanente y vertiginoso crecimiento turístico que aumenta la demanda energética. En consecuencia, es necesario incrementar la oferta energética; en este sentido, el trabajo tiene cuatro regiones que son las cuatro islas pobladas del archipiélago; Santa Cruz, San Cristóbal, Isabela y Floreana (ver Figura 2).

Figura 2
Infraestructura eléctrica por la Empresa Eléctrica de Galápagos



Datos

El trabajo usa diversas fuentes, la información sobre demandas eléctricas e infraestructura tecnológica existente proviene del SISDAT (ARCONER, 2023), que contiene datos históricos; las proyecciones de demanda del trabajo de la Fundación Bariloche (Fundación Bariloche FB, 2020). Los precios de los combustibles se consideran subsidios. Las proyecciones al año 2050 proceden del promedio histórico de subsidios del Banco Central del Ecuador del sector petrolero (BCE, 2023b).

Desarrollo del Modelo de Galápagos

Se definieron dos estaciones, la lluviosa y cálida que va desde diciembre hasta mayo, y la seca que va desde junio a noviembre. A continuación, se definieron las cuatro regiones; San Cristóbal (SC), Santa Cruz (STC), Isabela (IS), Floreana (FL). El horizonte de tiempo del modelo es de 2019 al 2050, y los primeros años se usan para calibrar al modelo. Los perfiles de los recursos provienen de las siguientes fuentes: El recurso viento obtiene datos de Ninja Wind, que se originan en el satélite de la NASA⁶, MERRA (Staffell and Pfenninger, 2016). Los datos de irradiancia global se adquieren del *System Advisor Model* (SAM, 2023), de la

6 The National Aeronautics and Space Administration.

National Renewable Energy Laboratory. Para estos perfiles se toma el año meteorológico que comprende promedios quince años o más. Las curvas de carga de cada isla forman parte de las demandas eléctricas. Finalmente, se tienen dos modos de operación debido a las baterías, uno que almacenan y en otro distribuyen energía.

Infraestructura Energética de Galápagos

La electricidad en las islas proviene de fuentes no renovables (84%), como las centrales térmicas de motores de combustión interna alimentadas con diésel (ver Tabla 1). Por tanto, para cambiar esta realidad se necesitan fuertes inversiones en energías limpias que con su desarrollo experimenten costos decrecientes progresivamente más accesibles para economías en desarrollo. Actualmente, el archipiélago cuenta con energía solar y eólica. Existen estudios del potencial energético de energía geotérmica en la isla Isabela y de olas en el sur de las islas (Portilla, Sosa, and Cavaleri 2013). Sin embargo, son tecnologías más costosas, especialmente la de las olas que aún se encuentra en desarrollo.

Tabla 1
Infraestructura eléctrica actual

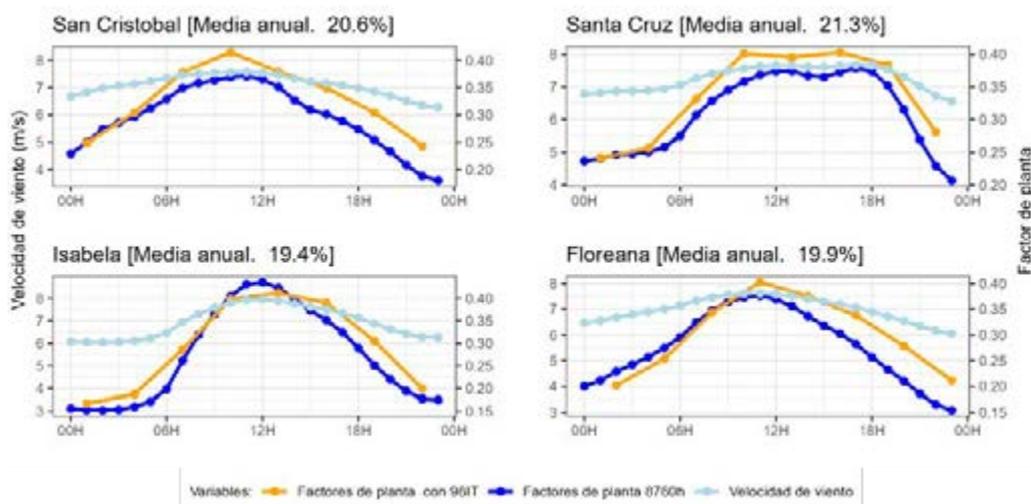
Central	Subtipo central	Isla	Puesta en operación	Vida operacional	Potencia nominal (kW)
Santa Cruz Solar Puerto Ayora	Solar	Santa Cruz/ Baltra	Diciembre de 2014	30 años	1.521
Baltra Solar	Solar	Santa Cruz/ Baltra	Enero de 2017	30 años	67
Santa Cruz Solar aislados	Solar	Santa Cruz/ Baltra	Diciembre de 2010	30 años	10
Baltra Eólico	Eólica	Santa Cruz/ Baltra	2014	25 años	2.250
Santa Cruz térmica	Térmica (MCI)	Santa Cruz/ Baltra	1990	15 años	14.810
San Cristóbal Solar Eolicosa	Solar	San Cristóbal	Julio de 2011	30 años	13
San Cristóbal Eólico	Eólica	San Cristóbal	2007	25 años	2.400
San Cristóbal térmica	Térmica (MCI)	San Cristóbal	1991	15 años	7.100
Isabela Solar aislados	Solar	Isabela	Enero de 2011	30 años	8
Isabela Solar	Solar	Isabela	Noviembre de 2018	30 años	953
Isabela Térmica	Térmica (MCI)	Isabela	2018	15 años	1.625
Floreana Perla Solar	Solar	Floreana	Septiembre de 2004	30 años	21
Floreana Solar aislados	Solar	Floreana	Septiembre de 2004	30 años	6
Floreana Térmica	Térmica (MCI)	Floreana	2010	15 años	440

Perfiles de los Recursos

Actualmente, el modelo OSeMOSYS no soporta perfiles horarios de los recursos; en consecuencia, fue necesario cambiar de perfiles de 8.760 horas al año a 96 intervalos de tiempo al año, que es lo que soporta el modelo. Para esta transformación se agruparon mediante promedios, ocho grupos de tres horas cada uno, por cada día del año, partiendo desde las cero horas; de esta forma se agregó los perfiles de los recursos al modelo. La Figura 3 ilustra dicha transformación

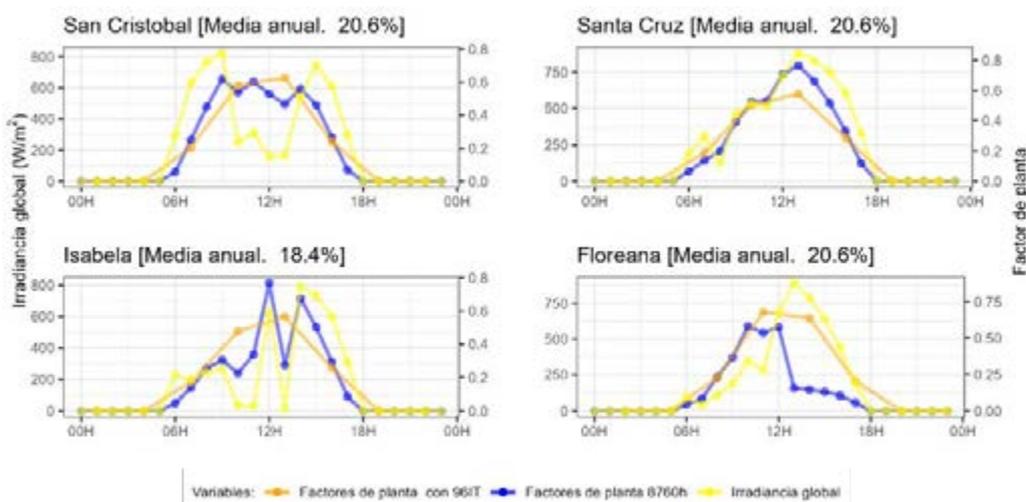
en un día de temporada seca, donde se puede observar la diferencia entre los factores de planta⁷ del recurso eólico con una resolución de 8.760 horas y los factores de planta con 96 intervalos de tiempo. Aunque se pierde algo de resolución, en general sigue la misma trayectoria del perfil del recurso (velocidad del viento). Cabe mencionar que al tratarse de un día de temporada seca, donde el recurso está en sus registros más altos, los factores de planta registran sus máximos valores; por lo tanto, en meses como enero dichos factores de planta serán más bajos; también se puede observar el promedio anual de la capacidad de las plantas en cada isla.

Figura 3
Perfil del viento en un día de la temporada seca - 15 de agosto



Para los perfiles solares se realizó el mismo procedimiento. Como resultado, se pueden observar factores de planta que llegan en promedio al 80% de la capacidad nominal en horas del mediodía, con promedios anuales de alrededor de 20% en las cuatro islas. Aunque al igual que en los perfiles eólicos sigue la trayectoria del recurso, se pierde la resolución de los picos (irradiancia global). Ver Figura 4.

Figura 4
Perfil solar en un día de la temporada seca - 15 de agosto



⁷ Es la relación de uso de la planta a su capacidad nominal, es decir, el porcentaje de horas al año en que la planta funciona a máxima capacidad.

Costos Empleados en las Tecnologías

Los costos de las dos tecnologías renovables empleadas en el modelo siguen decreciendo en el tiempo debido a que son tecnologías maduras. Los costos de las centrales termoeléctricas, así como su eficiencia, mejoran. Sin embargo, las termoeléctricas son alimentadas por combustibles fósiles, específicamente por diésel, lo que las hace vulnerables al mercado de materias primas y a las reservas con las que cuenta cada país, mismas que están decreciendo en todo el mundo. Por tanto, se espera que los combustibles sigan al alza (Ver Tabla 2). Además, se debe considerar que el planeta está en transición a energías renovables para alcanzar la carbono-neutralidad en términos de emisiones de GEI que permitan mitigar el cambio climático.

Tabla 2
Opciones de costos de electricidad

Opciones de planes de electricidad ¹	Costos de inversión (USD/kW)		Costos fijos de O&M (USD/kW/yr)		Costos Variables de O&M (USD/kW/yr)	
	2019	2050	2019	2050	2019	2050
Wind	2.114	1.158	35	17	0	0
Solar PV	1.740	960	17	10	0	0
Thermal (ICE)	1.200	1.200	20	20	1	1

¹ Fuente: Fundación Bariloche.

Resultados y Discusión

OSeMOSYS es un modelo de código abierto que permite su uso en diferentes regiones del mundo, así como también mejoras en el mismo, que cuenta con casos de estudio en más de cincuenta países alrededor del mundo⁸ (Detailed Model Documentation, 2023). Se mostrarán algunos casos de uso, incluido el caso de uso en las islas Galápagos con mayor detalle, esto con fines ilustrativos del alcance del modelo.

Casos de uso del Modelo OSeMOSYS en el Mundo

El trabajo de (De Moura, Legey, and Howells 2018) muestra las ventajas y desventajas en los procesos de interconexión energética con los países vecinos de Brasil (Argentina, Bolivia, Guayana, Paraguay y Perú). Sus escenarios fueron modelados con OSeMOSYS y comparados con los resultados del modelo OSeMOSYS SAMBA que incluye los países de Sudamérica. Como resultado, se obtuvo que Sudamérica tiene una alta dependencia de la generación hídrica. La comparación de escenarios SAMBA indica una reducción de hasta 23 GW en la expansión de la capacidad instalada brasileña y una adición de 5 GW en la capacidad instalada de los demás países de Sudamérica. Además, se mostró que países como Guyana y Bolivia, que tienen gran potencial hídrico pero poco financiamiento, pueden mejorar su infraestructura eléctrica con acuerdos energéticos con Brasil.

El trabajo de Quevedo and Moya (2022) presenta cuatro escenarios de expansión eléctrica con OSeMOSYS que han servido de insumo para los tomadores de decisiones. Se detallan las características de vulnerabilidad geográfica y su alta dependencia a las importaciones de combustibles fósiles. En consecuencia, establecieron como meta lo acordado por República

8 Las aplicaciones incluyen: EE. UU., México, Costa Rica, Nicaragua, Paraguay, Brasil, Bolivia, Ecuador, y toda Sudamérica (modelo SAMBA, país por país). UE27+ Reino Unido, Suiza y Noruega (país por país - modelo OSeMBE); todo el Continente africano (país por país - modelo TEMBA), Mauricio, Azerbaiyán, Irán, Pakistán, India, Vietnam, Nepal, Bután, Sri Lanka, Tailandia, Filipinas, Malasia, Mongolia, e Indonesia.

Dominicana en el Acuerdo de París, en donde, estimaron que para alcanzar dicho acuerdo sobre la reducción de un 25% de GEI para el año 2030, en comparación con las emisiones del año 2010, se necesita un total de inversión de USD 11.404 millones a una tasa de descuento de 4%.

Finalmente, el trabajo de Tehandao et al. (2023) modela la expansión eléctrica de Togo al año 2050. Se proponen tres escenarios; uno referencial que mantiene los niveles actuales de inversión en energía renovable, otro que propone cero emisiones de GEI al 2050 y el último que añade penalidades a las emisiones de GEI. Como resultado, el modelo propone energía solar e importaciones de combustibles como solución óptima a largo plazo, esto por encima del gas y de las hidroeléctricas, lo que implica una inversión diez veces mayor a la actual.

Caso de Uso en Galápagos

Escenarios de Expansión Eléctrica de Galápagos

Para el caso de las islas Galápagos se propusieron dos escenarios; el Referencial (REF) que sigue una trayectoria como la actual, con pequeñas inversiones en energías renovables y que al final del periodo tiene una alta proporción de energía no renovable proveniente de las centrales térmicas; y el escenario de descarbonización profunda llamado en este trabajo, Futuro Solar y Eólico (S&E). Ver Tabla 3)⁹.

Tabla 3
Escenarios a largo plazo de la expansión eléctrica en Galápagos

Escenarios	Descripción
Referencial (REF)	Energía térmica predominante (MCI) sin grandes inversiones en energías renovables (ER), con pequeñas adiciones en ER, alcanzando alrededor de 50 % de energía limpia al 2050 en el sector eléctrico en el archipiélago.
Futuro solar y eólico (S&E)	Fuerte inversión solar y eólica, alcanzando hasta un 75 % de energía limpia en 2030 y el 100 % hasta 2040 en el sector eléctrico.

El escenario Referencial (REF) busca seguir el mismo ritmo de inversión en energías renovables, es decir, sin mayor inversión en energías limpias; de esta forma se mantiene un mínimo de generación termoeléctrica. Como consecuencia, la matriz eléctrica no cambia significativamente, y las centrales térmicas (MCI) permanecen como la tecnología predominante en las islas (ver Figura 5). En la misma figura se puede observar el almacenamiento de energía eléctrica en naranja y la capacidad necesaria de almacenamiento hasta el año 2050. A pesar de que las centrales térmicas siguen dominando hasta el final del periodo modelado, este escenario alcanza alrededor de un 50% de electricidad limpia de fuentes renovables. También, se observa en el cuadrante de capacidad, que una vez cumplida la vida útil de las centrales térmicas, el modelo las reemplaza con centrales solares; lo mismo sucede con la tecnología eólica, que es reemplazada igualmente por la solar debido a que la tecnología eólica es más costosa. Al final del periodo, la mayor parte es energía térmica y solar. Con respecto a la capacidad de las baterías, es necesario alcanzar hasta 6 MW hasta el final del periodo, lo que permite cubrir la curva de demanda eléctrica.

En la Figura 6 se puede observar con mayor detalle la generación y capacidad eléctrica por isla; en Santa Cruz y San Cristóbal en el escenario REF se observa el mismo patrón que en el agregado de las cuatro islas, con un aumento tanto en la generación como en la capacidad solar.

⁹ Es importante señalar que se pueden probar los escenarios que sean necesarios de acuerdo con la información que se tenga disponible, en este caso se han mostrado solo dos escenarios para efectos ilustrativos del modelo OSeMOSYS.

Además, se observa un reemplazo de las centrales eólicas por las solares y una disminución de la capacidad de las centrales térmicas (MCI).

Por otro lado, en la Figura 7 se observa que Isabela alcanza a cubrir la demanda eléctrica con alrededor del 70% de energía solar y con mayor capacidad de baterías, de este modo disminuye la capacidad térmica en la isla. En lo que concierne a Floreana se observa que al final del periodo se tiene sobre el 50% de energía solar, así como una disminución importante de capacidad térmica en la isla.

Figura 5
Generación, capacidad y almacenamiento en el escenario referencial (REF)

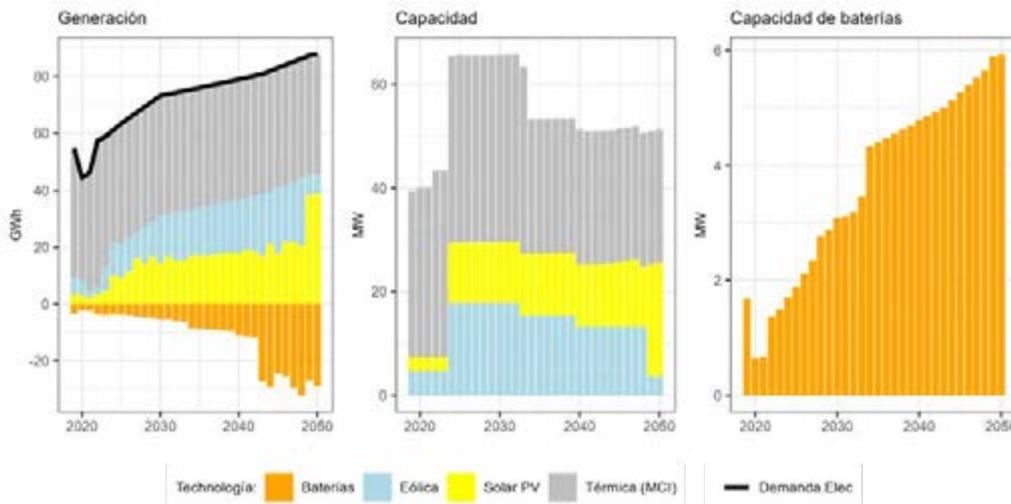
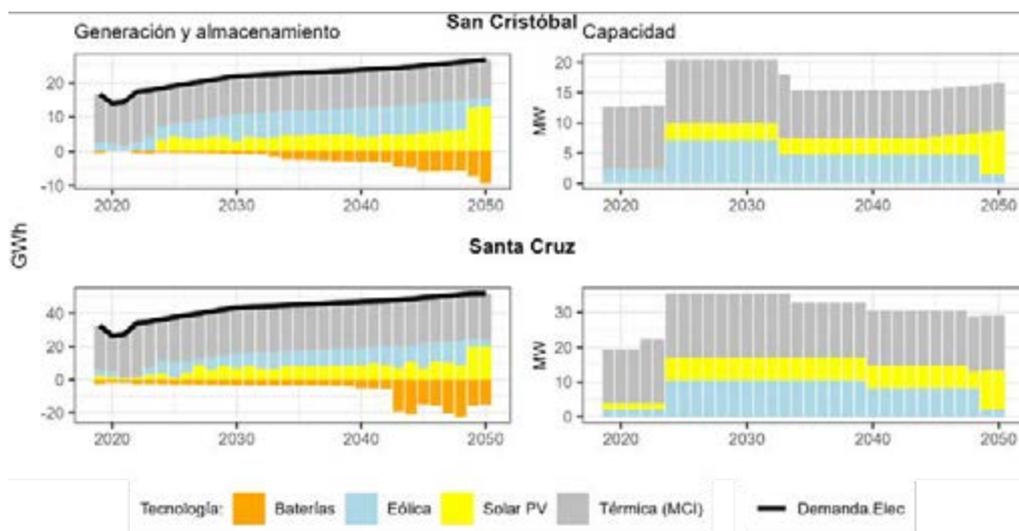


Figura 6
Generación, capacidad y almacenamiento de Santa Cruz y San Cristóbal en el escenario referencial (REF)



En el escenario de descarbonización profunda llamado Solar y Eólico (S&E), donde se asumen políticas agresivas de inversión en energía renovable, se alcanza alrededor de un 75% de energía limpia para el año 2030 y el 100% en el año 2040. Se observa que al final del periodo todas las centrales eólicas son reemplazadas por solares debido a su menor costo; por otro lado, la capacidad térmica se reduce casi en su totalidad y la capacidad de almacenamiento para cubrir

la demanda eléctrica sobrepasa los 10 MW (ver Figura 8). En consecuencia, las emisiones de GEI por generación eléctrica son eliminadas a partir del año 2040.

Figura 7

Generación, capacidad y almacenamiento de Isabela y Floreana en el escenario referencial (REF)

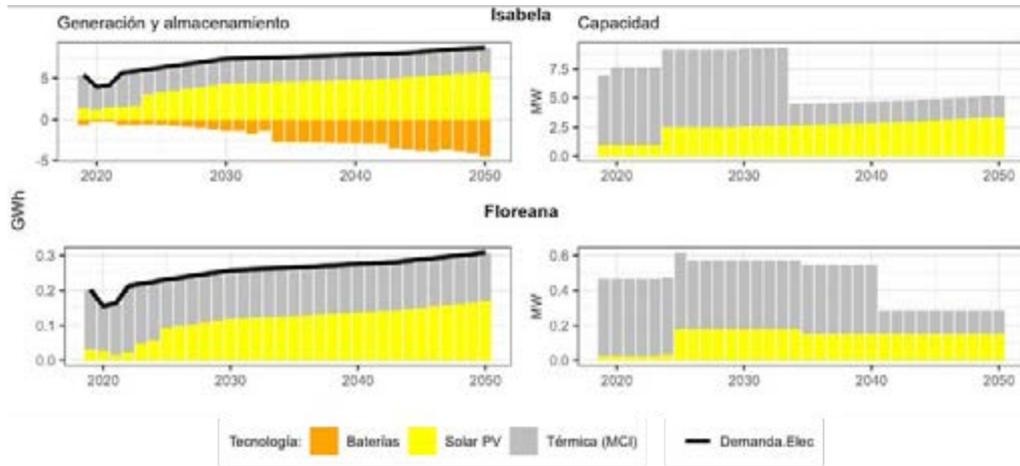
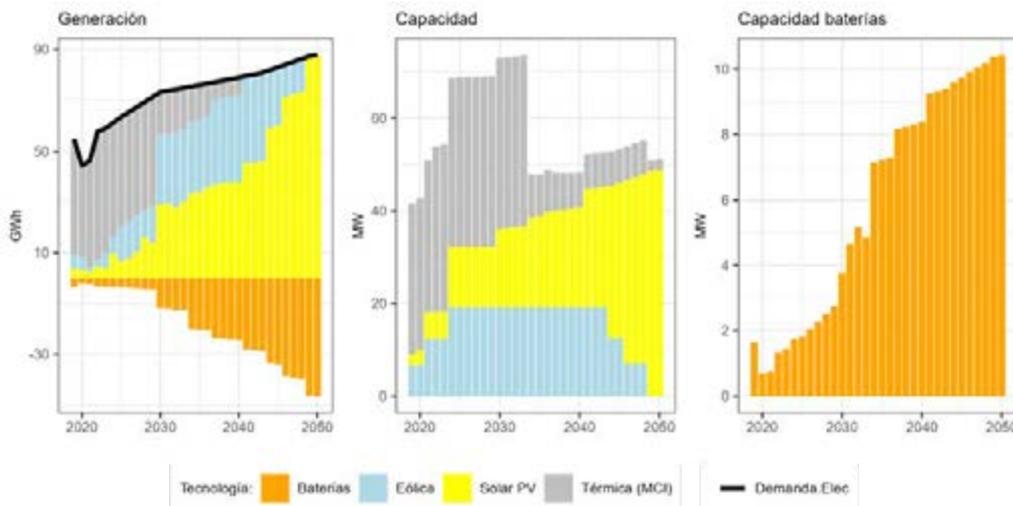


Figura 8

Generación, capacidad y almacenamiento en el escenario referencial Solar y Eólico



En la Figura 9 del escenario S&E se puede observar en detalle por isla; en San Cristóbal se alcanza el 100% de electricidad renovable en el año 2038, mientras que en Santa Cruz llega en el año 2040; esto debido a que es la isla más poblada del archipiélago. Con respecto a la expansión de capacidad, se tiene, al igual que en el escenario anterior, que ambas islas reemplazan tanto las centrales térmicas como las eólicas por solares.

Las islas con menor población Isabela y Floreana alcanzan más rápido el 100% de electricidad limpia, Isabela en el año 2034 y Floreana en el año 2026. Por otro lado, ambas islas llegan al final del periodo modelado con 100% de capacidad solar (ver Figura 10).

Figura 9
Generación, capacidad y almacenamiento de San Cristóbal y Santa Cruz en el escenario Solar y Eólico (S&E)

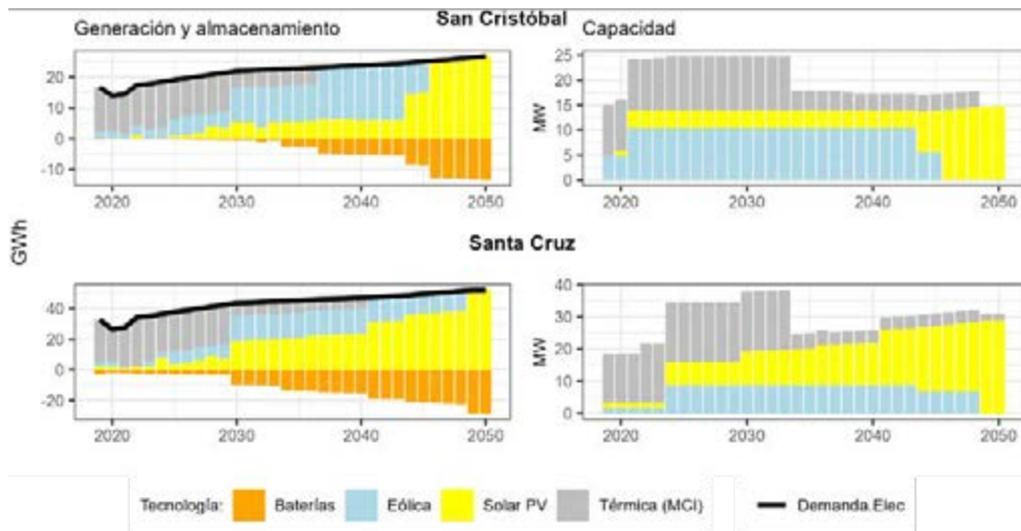
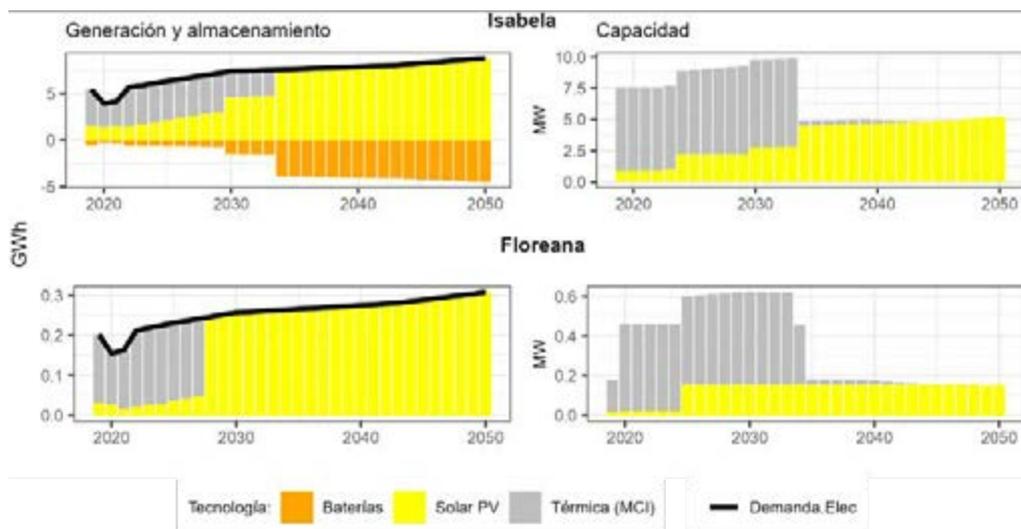


Figura 10
Generación, capacidad y almacenamiento de Isabela y Floreana en el escenario Solar y Eólico (S&E).

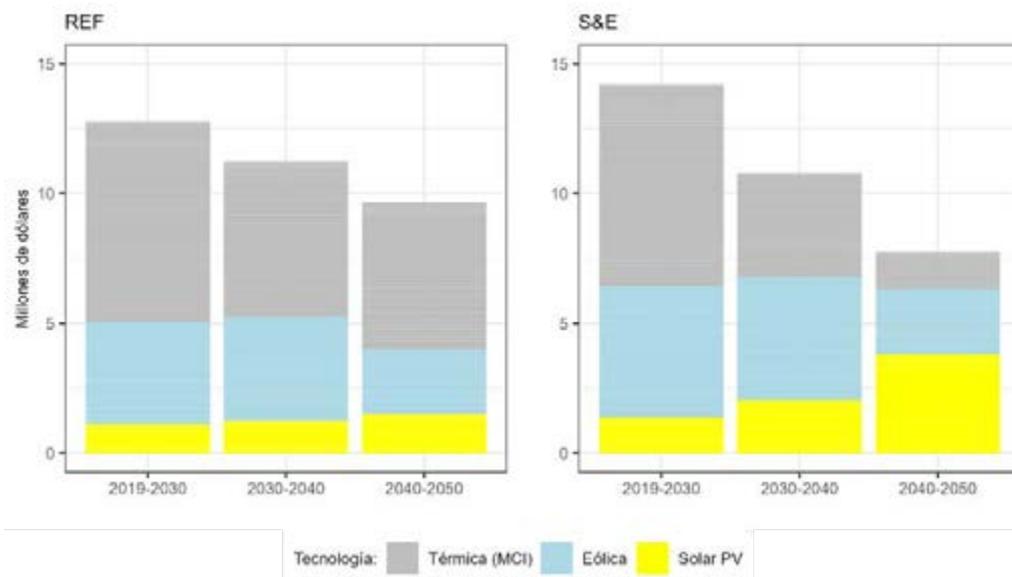


Costos de Inversión y de Operación y Mantenimiento

Para los costos de inversión, de operación y mantenimiento, tanto los fijos como los variables, se usa el valor presente neto en dólares del año 2021¹⁰. La Figura 11 muestra los costos de inversión, la primera década del periodo modelado el escenario Referencial (REF) es alrededor de USD 25 millones menor. Sin embargo, en las siguientes dos décadas del periodo, el escenario S&E tiene menos costos de inversión. En total, se tiene que el escenario REF sin mayor inversión es apenas USD 5 millones menos costoso que el escenario de descarbonización profunda del sector eléctrico en términos de inversión.

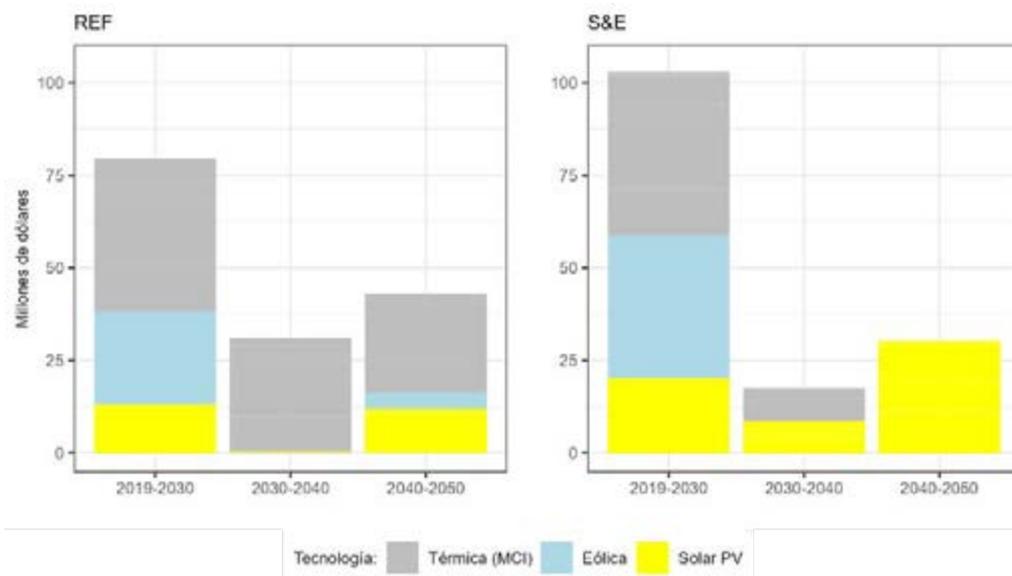
¹⁰ Año que se usa frecuentemente para traer valor presente debido a la apreciación y estabilidad que tuvo el dólar estadounidense.

Figura 11
Inversión por periodos y escenarios



En lo referente a los costos fijos de operación y mantenimiento, se tiene que el escenario REF en la primera década del modelo es menor que el escenario más ambicioso. En la segunda década es ligeramente menos costoso, y en la última es más costoso (ver Figura 12). Cabe recordar que los costos fijos usados tanto en la tecnología solar como en la eólica son más altos (ver Tabla 2).

Figura 12
Costos fijos de operación y mantenimiento por periodos y escenarios

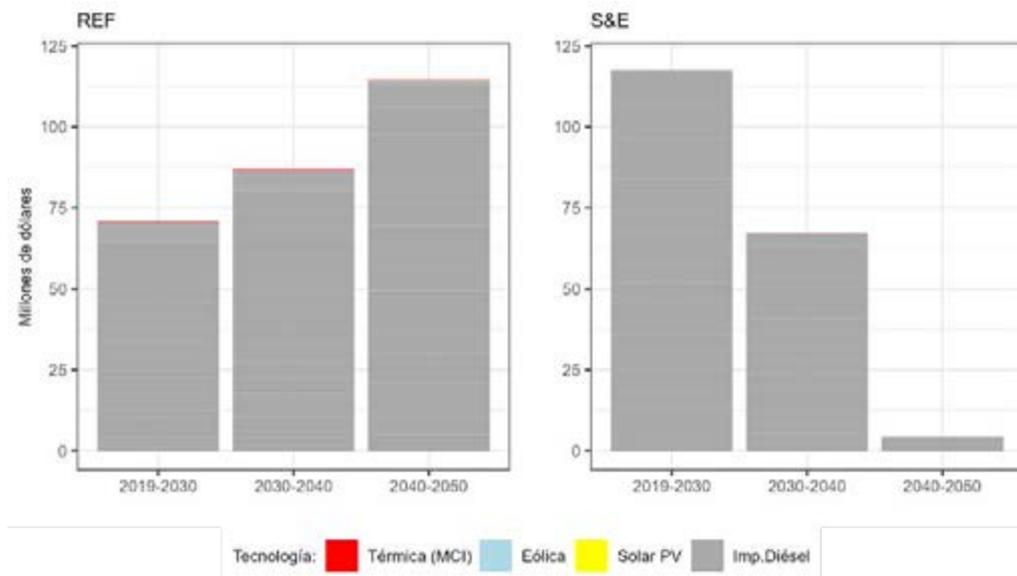


Las tecnologías renovables no presentan costos variables de operación y mantenimiento al usar recursos de libre disponibilidad, mientras que las centrales termoeléctricas tienen costos de 1,04 USD/MWh, a los que se les agregan los costos de importación de diésel que son altos. Como resultado, los costos variables del escenario REF tienden al alza a pesar de mantener subsidios. En la primera década del modelo los costos de ambos escenarios parten relativamente iguales, sin embargo, en los siguientes periodos son mucho menores. Al final del periodo, se tiene que en términos de costos variables el escenario REF es más del doble que el escenario

S&E, con una diferencia aproximada de USD 170 millones. La diferencia marginal en los costos de inversión y fijos entre el escenario REF y S&E no se compara con la diferencia en costos variables. Por tanto, apostar por la descarbonización del sector eléctrico es una inversión menos costosa que mantener las políticas energéticas actuales (ver Figura 13).

Figura 13

Costos de operación y mantenimiento variables por periodos y escenarios.



Conclusiones

El modelo OSeMOSYS de código abierto es flexible y versátil; fácil de usar; y tiene una curva corta de aprendizaje, lo que le permite desarrollarse rápidamente; con casos de uso en varias regiones del mundo y una importante comunidad internacional que hace posible su evolución. Una de las debilidades del modelo es que no soporta perfiles horarios de los recursos renovables. Esto puede ser debido a la cantidad de parámetros y variables que toma en cuenta. En consecuencia, el modelo requiere hacer una abreviación de los datos y se pierde algo de resolución. Por el contrario, el modelo trata con mayor detalle los costos, lo que permite ingresar costos de inversión, costos de operación y mantenimiento, fijos y variables, tasa de descuento de estos, costos de inversión de almacenamiento, costos de recuperación, entre otros.

Sobre los costos necesarios para la descarbonización del sector eléctrico, el modelo mostró que en términos de inversión el escenario referencial es apenas USD 5 millones menos costoso que el escenario de descarbonización profunda (S&E). En términos de costos fijos de operación y mantenimiento, se observa que el escenario REF es ligeramente menos costoso que el escenario S&E (USD millones).

Finalmente, los costos de O&M variables en el escenario S&E son considerablemente menores que los del escenario REF con un ahorro de USD 170 millones hasta el final del periodo en el 2050. En conclusión, no solo es posible la descarbonización del sector eléctrico, sino que también es una mejor inversión que continuar con las políticas actuales en el sector. Además de considerar lo anterior, se podría agregar lo que genera el turismo de conservación que valora la biodiversidad de las islas y su proyección de desarrollo con energías renovables, sumados los beneficios incalculables que se obtendrían al eliminar la posibilidad de derrames de combustibles en las costas del archipiélago. Por otro lado, es necesario recalcar que al tratarse

de un archipiélago megadiverso estas inversiones en energías alternativas se tienen que hacer con todas las precauciones debidas.

Por tanto, es sumamente importante que antes de comenzar cualquier proyecto de energías renovables se hagan los estudios necesarios a detalle de todos los posibles impactos ambientales que prioricen el uso de zonas ocupadas con asentamientos humanos. Una de las formas podría ser la energía no distribuida solar mediante paneles y baterías para el uso en viviendas y locales comerciales, lo que resulta clave en hábitats limitados como los de las islas Galápagos. Considero que de esta manera se puede incidir en la soberanía energética de las islas.

Es importante recalcar que la transición de energías no renovables a renovables no se enfoque enteramente en la reducción de emisiones de GEI, ya que por ejemplo las emisiones de GEI que producen las Galápagos son extremadamente marginales en la contribución global y debido a los permanentes vientos alisios, estas emisiones no se quedan en las islas. Por lo tanto, corresponde a toda la comunidad atacar los problemas de raíz y a la comunidad científica crear el conocimiento necesario para avanzar de forma acelerada hacia un futuro sostenible.

Para un próximo trabajo, se pueden desarrollar múltiples escenarios, como un escenario de expansión eléctrica con energía de olas, uno con energía geotérmica y/o uno con energía solar distribuida. Además, se pueden incluir todos los escenarios con y sin subsidios a los combustibles, los cuales representan una gran problemática en el país. Asimismo, se pueden proponer escenarios de expansión eléctrica que incluyan transporte terrestre eléctrico y/o marítimo, en donde la información disponible y la calidad de esta son las únicas limitantes del alcance del modelo.

Referencias

- ARCONER. 2023. SISDAT. <http://sisdat.controlrecursosyenergia.gob.ec/Login.aspx?ReturnUrl=%2fHome%2fFrame.aspx>.
- Astudillo, Fernando J. 2017. *Environmental Historical Archaeology of the Galápagos Islands: Paleoethnobotany of Hacienda El Progreso, 1870-1904*.
- Banco Central del Ecuador (BCE). 2023a. Cuenta Cantonales Sector Real. 2023. <https://contenido.bce.fin.ec/documentos/Estadisticas/SectorReal/CuentasCantonales/Indice.htm>
- Banco Central del Ecuador (BCE). Información Estadística. 2023. https://contenido.bce.fin.ec/documentos/Estadisticas/Hidrocarburos/indice_cifras.html
- Cannone, Carla, Lucy Allington, Lucy Allington, Philip Goynes, Cesar Valderrama, Vedran Kapor, et al. 2022. "clicSAND for OSeMOSYS: A User-Friendly Interface Using Open-Source Optimisation Software for Energy System Modelling Analysis." 2022. <https://assets.researchsquare.com/files/rs-1338761/v2/ffe64f16-9c9a-41e6-a298-5b0691391707.pdf?c=1663613864>.
- De Miras, Claude. 1995. *Las Islas Galápagos: Un Reto Económico: Tres Contradicciones básicas*. 513. Fundación Charles Darwin para las Islas Galápagos.
- De Moura, Gustavo Nikolaus Pinto, Luiz Fernando Loureiro Legey, and Mark Howells. 2018. "A Brazilian Perspective of Power Systems Integration Using OSeMOSYS SAMBA – South America Model Base – and the Bargaining Power of Neighbouring Countries: A Cooperative Games Approach." *Energy Policy* 115 (April): 470–85. <https://doi.org/10.1016/j.enpol.2018.01.045>.
- Epler, Bruce. 2007. Tourism, the Economy, Population Growth, and Conservation in Galapagos.

- Fundación Bariloche FB, BID. 2020. *Escenarios de demanda y oferta energética y opciones de política energética para el archipiélago de las islas Galápagos*. Vol. 4. 1. Quito.
- Gardumi, Francesco, Abhishek Shivakumar, Robbie Morrison, Constantinos Taliotis, Oliver Broad, Agnese Beltramo, Vignesh Sridharan, et al. 2018. "From the Development of an Open-Source Energy Modelling Tool to Its Application and the Creation of Communities of Practice: The Example of OSeMOSYS." *Energy Strategy Reviews* 20 (April): 209–28. <https://doi.org/10.1016/j.esr.2018.03.005>.
- Howells, Mark, Holger Rogner, Neil Strachan, Charles Heaps, Hillard Huntington, Socrates Kypreos, Alison Hughes, et al. 2011. "OSeMOSYS: The Open Source Energy Modeling System." *Energy Policy* 39 (10): 5850–70. <https://doi.org/10.1016/j.enpol.2011.06.033>.
- IEA. 2008. International Energy Agency, International Energy Workshop. 2008. https://www.internationalenergyworkshop.org/old/Workshop_2008.html.
- IIGE. 2018. Balance energético de la Provincia de Galápagos 2018.
- INEC. 2015. Censo de Población y Vivienda-Galápagos. 2015. <https://www.ecuadorencifras.gob.ec/censo-de-poblacion-y-vivienda-galapagos/>.
- Ministerio del ambiente, agua y transición ecológica (MAATE). 2023. INFORME ANUAL Ingreso de Visitantes a Las Áreas Protegidas de Galápagos. https://galapagos.gob.ec/wp-content/uploads/2023/02/INFORME_ANUAL_VISITANTES_2022_DUP.pdf
- Niet, T., A. Shivakumar, F. Gardumi, W. Usher, E. Williams, and M. Howells. 2021. Developing a Community of Practice Around an Open Source Energy Modelling Tool. *Energy Strategy Reviews* 35 (May): 100650. <https://doi.org/10.1016/j.esr.2021.100650>.
- OSeMOSYS. 2023. Structure of OSeMOSYS. 2023. <https://osemosys.readthedocs.io/en/latest/manual/Structure%20of%20OSeMOSYS.html>
- Pfenninger, Stefan, and Iain Staffell. 2016. "Long-Term Patterns of European PV Output Using 30 Years of Validated Hourly Reanalysis and Satellite Data." *Energy* 114 (November): 1251–65. <https://doi.org/10.1016/j.energy.2016.08.060>.
- PNG. 2023. Parque Nacional Galápagos | Sistema Nacional de Áreas Protegidas Del Ecuador. 2023. <http://areasprotegidas.ambiente.gob.ec/es/areas-protegidas/parque-nacional-gal%C3%A1pagos>
- PNUD. 2016. "Energía Verde para Galápagos Inagotable, Limpia y Segura." *Ministerio de Electricidad y Energía Renovable*, 22. https://www.undp.org/content/dam/ecuador/docs/documentos proyectos ambiente/pnud_ec REVISTA ENERGIA VERDE PARA GALAPAGOS
- Portilla, Jesus, Jeison Sosa, and Luigi Cavaleri. 2013. "Wave Energy Resources: Wave Climate and Exploitation." *Renewable Energy* 57 (September): 594–605. <https://doi.org/10.1016/j.renene.2013.02.032>.
- Quevedo, Jarrizon, and Idalberto Herrera Moya. 2022. "Modeling of the Dominican Republic Energy Systems with OSeMOSYS to Assess Alternative Scenarios for the Expansion of Renewable Energy Sources." *Energy Nexus* 6 (June): 100075. <https://doi.org/10.1016/j.nexus.2022.100075>.
- SAM. 2023. System Advisor Model - SAM. <https://sam.nrel.gov/>.
- Sampedro, Carolina, Francesco Pizzitutti, Diego Quiroga, Stephen J. Walsh, and Carlos F. Mena. 2020. "Food Supply System Dynamics in the Galapagos Islands: Agriculture, Livestock and Imports." *Renewable Agriculture and Food Systems* 35 (3): 234–48. <https://doi.org/10.1017/S1742170518000534>.
- Snell, HL, A Tye, CE Causton, and R Bensted-Smith. 2002. Current Status of and Threats to the Terrestrial Biodiversity of Galapagos. *A Biodiversity Vision for the Galapagos Islands*, 30–47.
- Staffell, Iain, and Stefan Pfenninger. 2016. "Using Bias-Corrected Reanalysis to Simulate Current and Future Wind Power Output." *Energy* 114 (November): 1224–39. <https://doi.org/10.1016/j.energy.2016.08.068>.

- Taylor, J. Edward, Jared Hardner, and Micki Stewart. 2009. Ecotourism and Economic Growth in the Galapagos: An Island Economy-Wide Analysis. *Environment and Development Economics* 14 (2): 139–62. <https://doi.org/10.1017/S1355770X08004646>.
- Tchandao, Ezzo-Wazam, Adekunlé Salami, Kodjo Koffi Mawugno, Amy Nabiliou, and Seydou Ouedraogo. 2023. Modelling the Optimal Electricity Mix for Togo by 2050 Using OSeMOSYS. *International Journal of Renewable Energy Development* 12 (February): 430–39. <https://doi.org/10.14710/ijred.2023.50104>.
- University, Open. 2023. Distance Learning Courses and Adult Education - the Open University 2023. <https://www.open.ac.uk/>.
- Walsh, Stephen J., and Carlos F. Mena. 2016. “Interactions of Social, Terrestrial, and Marine Sub-Systems in the Galapagos Islands, Ecuador.” *Proceedings of the National Academy of Sciences* 113 (51): 14536–43. <https://doi.org/10.1073/pnas.1604990113>.
- WWF. 2003. “Migración y Ambiente En Las Islas Galápagos ISBN 978-9978-58-129-2 - Libro. 2003. <https://isbn.cloud/9789978581292/migracion-y-ambiente-en-las-islas-galapagos/>.

Diseño de un Software para Apoyar tareas de Mantenimiento de Equipos de Laboratorio Clínico

Design of a Software to Support Clinical Laboratory Equipment Maintenance

Alex Fabián Guamán-Chango¹ <https://orcid.org/0009-0007-4705-3504>, Evelyn Cristina Faican-Jiménez¹ <https://orcid.org/0009-0006-6344-8082>, Ana Cecilia Villa-Parra^{1, 2} <https://orcid.org/0000-0002-7588-9372>

¹Universidad Politécnica Salesiana, Cuenca, Ecuador
aguamanc4@est.ups.edu.ec, efaicanj@est.ups.edu.ec

²Grupo de Investigación en Ingeniería Biomédica GIIB, Universidad Politécnica Salesiana, Cuenca, Ecuador
avilla@ups.edu.ec



Esta obra está bajo una licencia internacional
Creative Commons Atribución-NoComercial 4.0

Enviado: 2023/07/16

Aceptado: 2023/08/20

Publicado: 2023/10/15

Resumen

El mantenimiento de los equipos que se encuentran en laboratorios clínicos es requerido para garantizar su operación eficiente y prolongar su vida útil. Actualmente los procesos de mantenimiento se están gestionando a través de herramientas informáticas que deben ser desarrolladas a partir de los requerimientos de las instituciones y empresas. En este documento se presenta la descripción de un sistema para apoyar la gestión de las tareas de mantenimiento de equipos de laboratorio clínico de la empresa Proquímica de Ecuador. Se realizó un levantamiento de información en el Departamento de Mantenimiento de la empresa para contar con la información de soporte. Se diseñó y desarrolló un sistema, basado en una página web y una aplicación móvil, que permite a los técnicos y personal encargado del mantenimiento registrar de manera precisa y ordenada todas las actividades realizadas, incluyendo detalles como fechas, tareas y repuestos. La aplicación móvil brinda el acceso a los registros actualizados de mantenimiento de forma remota. Se realizó una evaluación de usabilidad del sistema con el cuestionario PSSUQ que es una herramienta que consta de 16 preguntas para evaluar la satisfacción de un usuario sobre una aplicación enfocándose en la facilidad de uso, la eficiencia

Sumario: Introducción, Materiales y Métodos, Resultados, Análisis de Resultados y Conclusiones.

Como citar: Guamán-Chango, A., Faican-Jiménez, E., Villa-Parra, A. C. (2023). Diseño de un Software para Apoyar tareas de Mantenimiento de Equipos de Laboratorio Clínico. *Revista Tecnológica - Espol*, 35(2), 94-108. Recuperado a partir de <http://www.rte.espol.edu.ec/index.php/tecnologica/article/view/1060>

y la calidad percibida. En la evaluación participaron cinco voluntarios y los resultados indican un alto nivel de satisfacción de la primera versión del sistema y de los recursos que brinda para realizar las tareas de mantenimiento.

Palabras clave: mantenimiento, equipo clínico, aplicación móvil, página web.

Abstract

Periodic maintenance of equipment found in clinical laboratories is necessary to ensure its efficient operation and extend its lifespan. Currently, maintenance processes are managed through computer tools, which need to be developed based on the requirements of institutions and companies. This work presents the description of a system to support the maintenance tasks of clinical laboratory equipment at the company Proquímica, Ecuador. Information was gathered from the Maintenance Department of the company to provide supporting information for the maintenance of clinical equipment. In addition, a system was designed and developed, which includes a website and a mobile application. This allows technicians and maintenance personnel to accurately and orderly record all activities performed, including details such as dates, tasks, and spare parts. The mobile application provides remote access to up-to-date maintenance records. A usability evaluation of the system was conducted using the PSSUQ questionnaire, which consists of 16 questions to assess user satisfaction with an application, focusing on ease of use, efficiency, and perceived quality. Five volunteers participated in the evaluation, and the results indicate a high level of satisfaction with the system and the resources it provides for performing maintenance tasks.

Keywords: maintenance, clinical equipment, mobile application, website.

Introducción

El mantenimiento de equipo médico en ambientes hospitalarios y de laboratorio clínico, es trascendental para la operación eficiente de estos espacios y para garantizar la precisión de los resultados. Realizar el mantenimiento de los equipos permite prolongar su vida útil y garantizar la seguridad del personal que opera en el laboratorio e instituciones prestadoras de servicios de salud (Iadanza et al., 2019). La gestión eficaz del mantenimiento de equipos médicos además influye directamente en la calidad de la atención y es esencial para proporcionar buenos servicios de salud, reducir la mortalidad y ahorrar recursos (Badnjevic, 2023; Bahreini et al., 2019; Márquez, 2023).

Considerando que los equipos que se emplean para realizar pruebas y análisis son la base para el diagnóstico y tratamiento de las personas, estos deben contar con un plan de mantenimiento que asegure su trabajo en óptimas condiciones, con evidencias de su calibración y funcionalidad (Zamzam et al., 2021a). En este sentido, las empresas proveedoras de equipos médicos ofrecen servicios de mantenimiento y control de calidad, que deben proyectarse durante toda la vida útil del equipo. Para su puesta en marcha, el mantenimiento exige una planificación eficiente, en la cual se integren de manera efectiva los responsables del servicio técnico y los clientes como instituciones públicas y privadas. Debilidades en la correcta administración del recurso humano y en la planificación, programación y control del mantenimiento, provocan que las instituciones prestadoras de servicios de salud registren evaluaciones de la gestión del mantenimiento no satisfactorias (Cuzco et al., 2019), que exigen mejoras inmediatas en los departamentos responsables.

El mantenimiento contempla conocer el manual de usuario de los equipos, para saber en detalle cómo se utiliza correctamente el dispositivo en la práctica clínica, además se requiere llevar un registro de las fechas de adquisición e inspecciones, el diseño de planes de mantenimiento preventivo y correctivo incluyendo la capacitación en los procesos que garanticen la calibración y ajustes de los equipos (Pérez et al., 2021; WHO, 2012).

La evaluación del estado de los equipos médicos y la gestión de su ciclo de vida son actividades importantes durante el mantenimiento del equipo, que es clave para aumentar la disponibilidad, el rendimiento y la seguridad (Zamzam et al., 2021a). De acuerdo con Bahreini et al. (2019), entre los factores que tiene influencia directa en los sistemas de gestión de mantenimiento eficientes están: los recursos, el control de calidad, la educación y el diseño e implementación de los planes de mantenimiento. Márquez (2023) indica que, para garantizar la calidad de los dispositivos médicos en la prestación de atención médica, se deben desarrollar listas de verificación que identifiquen el estado de desempeño del mantenimiento de los equipos.

Por lo mencionado, llevar un registro detallado de todas las actividades de mantenimiento realizadas es importante para contar con la información completa para la verificación de la gestión de mantenimiento. En muchos casos, la documentación básica se lleva en registros en papel, pero el uso de sistemas computarizados de gestión del mantenimiento facilita y optimiza el registro y el seguimiento del desempeño en entornos médicos (WHO, 2012), con herramientas informáticas, por ejemplo, para la gestión de mantenimiento asistido por ordenador conocidas como GMAO (Valenzuela, 2023; Ávila et al., 2022). Existen además sistemas para la gestión de tareas de mantenimiento y que son aplicadas y evaluadas en centros prestadores de salud, como el sistema computarizado de gestión de mantenimiento CMMS por sus siglas en inglés “Computerized Maintenance Management System” y el sistema para gestión de activos EAM por sus siglas en inglés “Enterprise Asset Management” (Ávila et al., 2022; Cuzco, et al., 2019). Estos recursos cuentan con funcionalidades que incluyen la gestión de inventario, la programación de visitas para el mantenimiento, la creación del historial de mantenimiento y de bases de datos tanto de proveedores como de clientes, la generación de informes, alertas y recordatorios.

Herramientas avanzadas, mediante internet de las cosas y modelos predictivos, pueden evaluar los resultados del mantenimiento y proyectar fallos y problemas en los equipos, permitiendo realizar intervenciones preventivas oportunas y eficientes (Shamayleh et al., 2020; Silva et al., 2021; Zamzam et al., 2021b).

Como indica Badnjevic (2023), para optimizar los programas de mantenimiento y reducir su costo total, las estructuras de gestión hospitalaria buscan constantemente soluciones para extender el tiempo de operación de los equipos, en los niveles técnicos y de seguridad. En el campo del laboratorio clínico, la tecnología desempeña un papel crucial para mejorar la eficiencia, la precisión y la gestión de los procesos. El software y aplicaciones móviles diseñadas para estos equipos han cambiado la forma en la que se llevan a cabo las pruebas y se gestionan los resultados, sin embargo, son pocas las opciones que realizan una interpretación de los datos generados para identificar el estado de los equipos y garantizar su vida útil. Existe el requerimiento entonces de integrar los recursos de software para que los datos generados por los equipos sean procesados para mejorar las tareas de mantenimiento (Iadanza et al., 2019).

Se debe considerar además que, el emplear un software para gestionar actividades de mantenimiento plantea a las empresas desafíos como una inversión por costo de licencias, actualizaciones y soporte técnico. Se debe considerar el nivel de complejidad en la implementación

de las funcionalidades de un software para estas aplicaciones, que requiere ser adaptado a las necesidades de las instituciones y ofrecer la capacitación a su personal, para asegurar que se adopte y utilice de manera efectiva en las rutinas de trabajo (Badnjevic, 2023).

Por esta razón, el diseño de un software para mantenimiento requiere definir los requisitos y necesidades de la institución o empresa para ajustar las funcionalidades de la herramienta informática al flujo de trabajo y procesos. Adicionalmente, es necesario que se evalúen las opciones de software comparando características, capacidades, costo y soporte ofrecidos, para elegir la que se ajuste mejor. La implementación gradual brinda la posibilidad de iniciar con una herramienta que ayude a la gestión con funcionalidades básicas que puedan migrar a procesos más complejos a la par con la capacitación del personal, que debe ingresar la información y gestionarla de manera correcta. Es importante en este proceso evaluar continuamente el rendimiento para identificar los ajustes que mejoren la interacción con el software (Kumar et al., 2021).

En este contexto, se presenta el desarrollo de un sistema para gestionar las tareas de mantenimiento de equipos de laboratorio clínico para la empresa ecuatoriana Proquímica. Se parte con la descripción del diseño del software adaptado a la empresa, que permita contar con un recurso eficiente y útil en la mejora de los procesos de mantenimiento. Posteriormente, se describe la metodología aplicada para la recolección de la información, para el desarrollo del software, de la interfaz y las pruebas. Se presentan finalmente los resultados de la evaluación de usabilidad realizada por los colaboradores de la empresa y las conclusiones.

Materiales y Métodos

Recopilación de información

El sistema desarrollado fue alimentado con la información de los procedimientos para mantenimiento seguidos por el Departamento de Mantenimiento de la empresa Proquímica (DMPQ), ubicada en la ciudad de Cuenca, Ecuador, con 25 años de experiencia como distribuidor de equipos para laboratorio clínico, control de calidad y prestador de servicios de mantenimiento de este tipo de equipos de complejidad baja y media.

Al realizar el levantamiento de información en la empresa, se consideraron los aspectos de funcionalidad, requisitos técnicos, normativas, procesos, configuración y protocolos de manipulación de muestras y seguridad. Se identificaron estos conceptos para proponer los requisitos del sistema que apoye a la gestión de tareas de mantenimiento de los equipos de laboratorio clínico que la empresa realiza posventa.

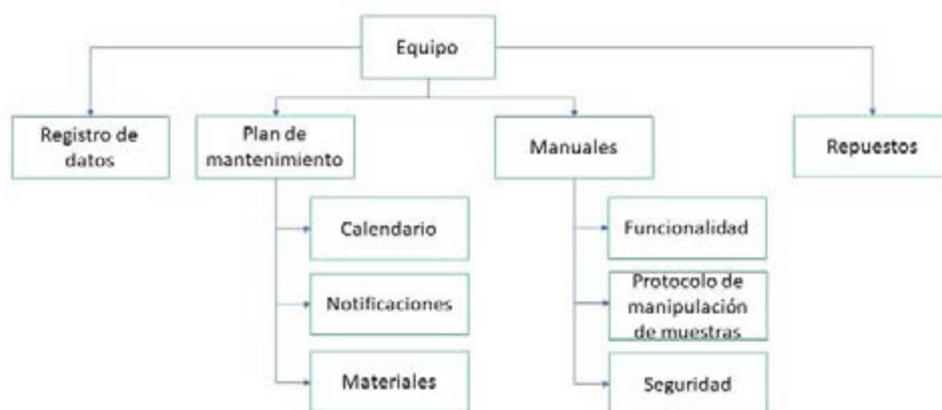
En el diagrama de la Figura 1 se muestra la propuesta de información relevante para alimentar el sistema que brinde soporte al mantenimiento de estos equipos de laboratorio clínico.

Para el registro del equipo son necesarios los datos técnicos que deberán ser ingresados al sistema para que los responsables de realizar el mantenimiento puedan crear una ficha del equipo con: nombre del equipo, modelo, número de serie, fecha de adquisición y responsable de mantenimiento. Esta información es considerada relevante para evaluar la confiabilidad de dispositivos médicos (Zamzam et al., 2021a).

El plan de mantenimiento requiere incluir un calendario para la programación de las fechas de mantenimiento y de las tareas de inspección.

Figura 1

Diagrama de bloques con la descripción de la información requerida para la gestión de mantenimiento de un equipo de laboratorio clínico



Con las notificaciones se busca implementar un sistema con indicaciones y alertas para la ejecución de tareas de mantenimiento para los equipos.

La sección de materiales busca que los técnicos cuenten con los recursos necesarios para realizar los servicios de mantenimiento a tiempo.

El acceso a los manuales y documentación técnica busca facilitar el acceso a guías de mantenimiento, especificaciones técnicas y documentación relevante para el proceso de mantenimiento. Se incluye: la descripción de las funciones y configuraciones de los equipos dependiendo de los modelos, las normas y estándares aplicables, los requisitos de verificación, calibración, mantenimiento y seguridad. Los equipos cuentan con la descripción de protocolos especiales para una manipulación adecuada de muestras. Esta información es considerada en procesos de evaluación de la gestión de mantenimiento (Cuzco et al., 2019).

La información de repuestos busca facilitar el control de existencias de los insumos necesarios para realizar el mantenimiento.

Se realizó una serie de visitas, en las cuales se conocieron los detalles de la Figura 1 y las tareas de mantenimiento llevadas a cabo para los siguientes equipos de laboratorio clínico, que son los más demandados en la empresa: centrífuga, microscopio, baño maría e incubadora. Se revisó, con previa autorización de la empresa, la documentación de los equipos y de las tareas de mantenimiento, que se encontraba registrada en hojas de papel y que fue digitalizada. El proceso de recolección, ordenamiento y análisis de toda la información se realizó en 30 días en jornadas de 8 horas de trabajo diarias.

La investigación incluyó el levantamiento de información en el DMPQ para conocer sobre el tipo de mantenimiento ofrecido por la empresa (preventivo, correctivo y predictivo), la frecuencia y la duración de las intervenciones, los materiales e insumos para llevar a cabo las actividades de mantenimiento y los procedimientos realizados para el mantenimiento de los equipos: centrífuga, microscopio, baño maría e incubadora.

En la tabla 1 se encuentra la ficha con la información solicitada al DMPQ.

Tabla 1
 Ficha para el registro de las actividades de mantenimiento llevadas a cabo en la empresa Proquímica

Actividades de Mantenimiento						
Equipo	Tipo de Mantenimiento		Frecuencia	Duración	Materiales e Insumos	Procedimiento
	Preventivo	<input type="checkbox"/>				
	Correctivo	<input type="checkbox"/>				
	Predictivo	<input type="checkbox"/>				

Con base en la información recopilada se creó un plan de mantenimiento preventivo y correctivo para estos equipos.

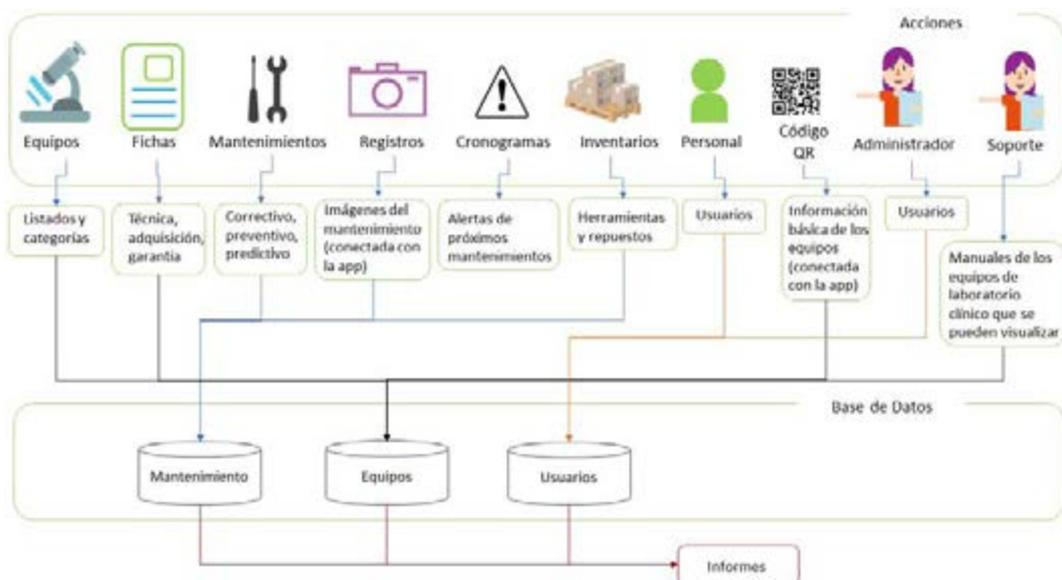
Al consultar a los ingenieros del DMPQ si existía la predisposición para incluir un sistema de apoyo a la gestión del mantenimiento de los equipos y utilizarlo como herramienta en su trabajo, se recibió una respuesta afirmativa. El DMPQ planteó como requerimiento que sea una herramienta que permita automatizar la gestión de servicios, el seguimiento de órdenes de trabajo y el registro de datos técnicos.

Se verificó que la infraestructura tecnológica del DMPQ, basada en computadoras portátiles y teléfonos inteligentes, permitía la implementación de un sistema para gestionar las tareas de mantenimiento.

Sistema para Gestionar el Mantenimiento de Equipo de Laboratorio Clínico (SGM-ELC)

En la Figura 2 se presenta la estructura del sistema SGM-ELC para gestionar el mantenimiento de equipo de laboratorio clínico, que incluye una página web, una aplicación móvil y una base de datos.

Figura 2
 Diagrama de flujo del sistema que muestra la interconexión de las acciones e información del sistema SGM-ELC para la gestión del mantenimiento de equipos de laboratorio clínico



El software incorpora información clara de los equipos con las secciones: equipo, categoría, fichas y código QR, por sus siglas en inglés “Quick Response”.

Para ayudar en el seguimiento de las actividades de mantenimiento preventivo y correctivo de los equipos se incorporaron las secciones: mantenimiento, registro, cronogramas, inventarios y soporte. Además, se cuenta con información del personal y sus responsabilidades con las secciones: personal y administrador. Esto brinda una base que apoya las tareas de mantenimiento en el momento adecuado y de manera sistemática, para prevenir averías y reducir el tiempo de inactividad no planificado.

Para almacenar y administrar la información relacionada con el mantenimiento de equipos de laboratorio clínico, SGM-ELC crea bases de datos con la información de mantenimiento, equipos y usuarios, de tal manera que se puedan almacenar y recuperar los datos necesarios para la gestión del mantenimiento. La información puede exportarse como documentos con extensión pdf.

Para el desarrollo de la página web se utilizó Visual Code. El diseño visual de la interfaz de la página se realizó con una plantilla de la biblioteca Bootstrap, como una estructura conceptual y una base adecuada para el desarrollo de la primera versión del sistema.

Considerando que los trabajadores de la empresa usarían por primera vez un software para fines de mantenimiento de equipos se buscó que sea una herramienta que permita fácil interacción.

Durante el proceso de desarrollo, se utilizaron herramientas como Eclipse, Visual Studio o IntelliJ IDEA, que ofrecen un entorno completo para escribir, depurar y probar el software. Para gestionar el código fuente y colaborar en equipo se utilizó la herramienta de control de versiones Git.

Para realizar la comunicación y transferencia de datos entre diferentes sistemas con el software se utilizó la interfaz API, por sus siglas en inglés “Application Programming Interfaces” y servicios web. Esta interfaz permite integrar el software de mantenimiento de equipos de laboratorio clínico con otros sistemas o acceder a información externa relevante.

El sistema fue diseñado para contar con una aplicación móvil desarrollada en el software Android para permitir que los responsables de mantenimiento estén conectados con la página web y puedan gestionar información de las siguientes secciones: fichas, mantenimientos, registros, escanear código QR y soporte.

El código QR es el medio por el cual se puede acceder a la información básica de los equipos mediante la aplicación. La información de evidencias de las condiciones del equipo antes y después del mantenimiento puede ser revisada con la aplicación móvil para facilitar el envío de información por otros medios de comunicación como correo electrónico cuando se requiera.

Para crear las interfaces interactivas de usuario de la aplicación, se utilizaron las herramientas HTML, CSS y frameworks y para la gestión de las bases de datos del sistema se utilizaron MySQL, PostgreSQL, Oracle y SQL Server.

La interacción entre la página web y la aplicación se basa en las conexiones con la base de datos. La aplicación, después de establecer la conexión MySQL permite acceder a una base de datos con fotos para subir y descargar evidencias de los mantenimientos realizados. Esta información forma parte de los reportes de mantenimiento e información de los equipos.

El sistema cuenta con encriptación de datos en tránsito y en reposo, mediante protocolos HTTPS que permiten la comunicación entre el cliente y el servidor. La encriptación en reposo implica que los datos almacenados en los servidores y en las bases de datos también estén encriptados para controlar el acceso no autorizado. Para acceder se exige la autenticación de los usuarios con una contraseña. Se implementaron niveles de autorización para asegurar el acceso a los datos y a las funciones de ingreso de información del equipo de laboratorio clínico.

Realizada la programación del sistema, se incorporó la información que brindó la empresa Proquímica, considerando subir documentos debidamente referenciados y organizados, tomando en cuenta que mucha de esta información fue recopilada directamente de los trabajadores de la empresa.

Como el sistema fue concebido de manera modular, es posible incorporar nuevos bloques que amplíen su funcionalidad.

Pruebas de usabilidad del sistema

Las pruebas de usabilidad del sistema SGM-ELC fueron realizadas por 5 voluntarios, 4 hombres y 1 mujer ($37,7 \pm 14$ años), ingenieros del DMPQ.

Los participantes recibieron una explicación-capacitación sobre todo el proceso de uso del sistema.

Se solicitó a los participantes realizar las siguientes actividades: 1) instalar la aplicación móvil. 2) ingresar la información de un equipo en todas las categorías indicadas en la Figura 2 en la página web. 3) generar los informes de las bases de datos del sistema en la página web y con la aplicación móvil. 4) verificar si los informes se encuentran completos y en el formato correcto; y 5) llenar el formulario PSSUQ, por sus siglas en inglés Post-Study System Usability Questionnaire, de la Tabla 2.

El PSSUQ es una herramienta utilizada para evaluar la satisfacción del usuario sobre un sistema o una aplicación. Este cuestionario de 16 preguntas, se enfoca en la facilidad de uso, la eficiencia y la calidad percibida del sistema y se administra a los usuarios después de que han interactuado con un sistema o producto (Hajesmaeel-Gohari et al., 2022). Los valores más cercanos a 1 representan una mejor usabilidad, el valor neutro es 4 y los más cercanos a 7 representan una menor usabilidad.

En el PSSUQ, las puntuaciones más bajas significan una mejor evaluación, lo que indica un mayor nivel de usabilidad y en la Tabla 2 se muestran las preguntas del cuestionario y las escalas para su evaluación (Villa-Parra et al., 2022).

Tabla 2

Preguntas del cuestionario PSSUQ y sus escalas aplicado para la evaluación del sistema para el mantenimiento de los equipos CEN, MIC, BM e INC

		Mayor usabilidad			Neutral			Menor usabilidad
#	Pregunta	1	2	3	4	5	6	7
1	Estoy satisfecho/a con este sistema							
2	Fue simple usar este sistema							
3	Soy capaz de completar mi trabajo rápidamente utilizando este sistema							
4	Me sentí cómodo usando este sistema							
5	Fue fácil aprender a usar este sistema							
6	Creo que me volví experto rápidamente utilizando este sistema							
7	Creo que podría volverme productivo rápidamente usando este sistema							
8	El sistema dio mensajes de error que claramente me dijeron cómo solucionar problemas							
9	Cada vez que cometía un error al usar el sistema, podía recuperarme fácil y rápidamente							
10	La información (como ayuda en línea, mensajes en pantalla y otra documentación) proporcionada con este sistema fue clara							
11	Fue fácil encontrar la información que necesitaba							
12	La organización de la información del sistema en la pantalla fue clara							
13	La interfaz de este sistema era agradable							
14	El sistema tuvo todas las herramientas que esperaba que tuviera							
15	Este sistema tiene todas las funciones y capacidades que espero que tenga							
16	En general, estoy satisfecho con este sistema							

Resultados

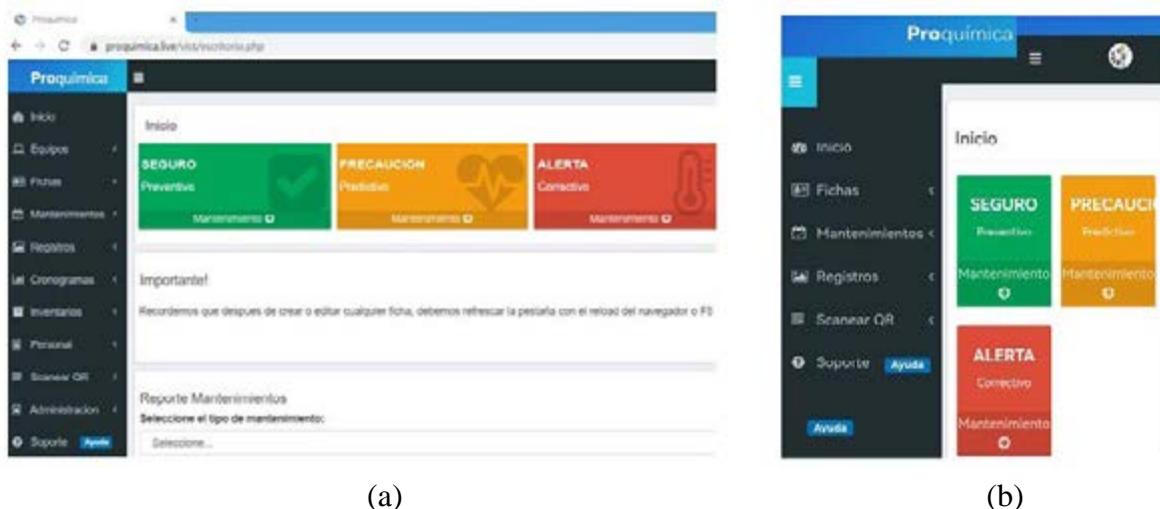
En la Figura 3 se muestran las interfaces para realizar el acceso al sistema (login), tanto para la página web mediante una PC y a la aplicación móvil mediante un teléfono inteligente. Se deben introducir los datos de usuario y la contraseña.

Figura 3
Autenticación del sistema SGM-ELC (a) página web; (b) aplicación móvil



En las páginas principales de la página web y de la App que se muestran en la Figura 4, se despliegan noticias que son incluidas por el administrador y alertas de mantenimiento que son generadas de manera automática con base en el cronograma registrado en el calendario de trabajo.

Figura 4
Interfaz del sistema SGM-ELC (a) página web; (b) aplicación móvil



Con la opción de lista de equipos, en la página web es posible imprimir en formato PDF, por sus siglas en inglés “Portable Document Format”, un informe con los datos del equipo (equipo, código, fecha), fichas (técnica, de garantía, de adquisición), y registros del mantenimiento (predictivo, correctivo, preventivo). En la Figura 5, se muestra como ejemplo la información registrada de la centrífuga.

Figura 5

Información registrada de la centrífuga que se visualiza en un documento en formato PDF

Equipos **Reporte**

Nombre(*):
CENTRIFUGA

Stock
1

Imagen:
Seleccionar archivo Sin archivos seleccionados

Categoría(*):
Maquinaria Principal

Descripción
LABORATORIO CLÍNICO

Código:
C1001

Generar **Imprimir**

C1001

Guardar **Cancelar**

Se puede observar que se requiere el ingreso de los datos como el nombre del equipo, el número de existencias, la descripción del equipo, una imagen para especificar el tipo de equipo y el código de identificación que tendrá en el sistema.

Con la aplicación se puede acceder a la información de los equipos y se verificó que es posible acceder a los reportes fotográficos de los mantenimientos. Mediante el código QR se puede revisar la información completa que apoye el proceso de mantenimiento. En la sección de soporte se puede acceder a los documentos con la guía para realizar el mantenimiento de los equipos centrífuga, microscopio, baño maría e incubadora.

En la Figura 6 se muestra la interacción de los participantes en las pruebas con la página web y con la aplicación del sistema.

Figura 6

Participantes realizando las pruebas de usabilidad del sistema SGM-ELC en la empresa Proquímica (a) interacción con la página web; (b) acceso a la aplicación móvil



(a)

(b)

En el marco de la capacitación para la instalación y uso del sistema, se respondieron a las dudas que surgieron en el proceso.

En la tabla 3 se presentan los resultados de la aplicación del cuestionario PSSUQ a los participantes de las pruebas, en donde 1 representa una mayor usabilidad, 4 es un valor neutro y los valores cercanos a 7 representan una menor usabilidad.

Tabla 3
Resultados del cuestionario PSSUQ aplicado a los participantes de las pruebas del sistema SGM-ELC

Preguntas																
Voluntario	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
2	1	2	1	1	2	2	2	1	1	1	1	1	1	1	1	1
3	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
4	2	1	1	2	2	2	1	1	2	2	1	1	1	1	2	1
5	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1

Análisis de Resultados

En la recolección, organización y análisis de la información fue importante contar con la ayuda del personal de la empresa, quienes sirvieron de guía para proporcionar la información más importante al sistema.

El levantamiento de información, basada en la experiencia de los ingenieros que laboran en la empresa Proquímica, permitió que los planes de mantenimientos puedan ser ingresados al sistema para que futuros trabajadores puedan tener acceso a información relevante de la empresa.

Durante el desarrollo de las pruebas no existieron problemas con la instalación de la aplicación móvil y con el manejo de la página web en los computadores de la empresa. Durante la capacitación para la instalación y el uso se realizaron preguntas sobre la sincronización de la aplicación móvil con la página web. Se destacó que se brindaron explicaciones claras sobre la instalación y el uso del sistema.

Con base en los resultados del cuestionario PSSUQ presentados en la Tabla 3, entre las preguntas mejor evaluadas, están las preguntas 3, 8, 11, 12, 13 y 14, las cuales recibieron la respuesta de una mayor usabilidad por parte del 100% de los participantes. La pregunta 3 resalta el apoyo que puede brindar el sistema a las tareas de mantenimiento de los equipos. Las preguntas 8, 11, 12, 13 y 14, indican que los voluntarios perciben que el sistema posee la información necesaria y brinda una guía útil para interactuar con el mismo.

Las preguntas 5 y 6, relacionadas al nivel de dificultad para aprender a usar este sistema, presentan una menor puntuación. El 60% de los participantes manifestaron plena satisfacción. Cabe mencionar, que los participantes de mayor edad del grupo (40%) requirieron de mayor tiempo para acceder al sistema y no manifestaron estar totalmente satisfechos con el sistema. Por estos resultados, es necesario brindar un proceso de capacitación que profundice en la explicación del manejo de recursos informáticos.

En las preguntas 1, 2, 4, 7, 9, 10 y 15, al menos un voluntario no evaluó con plena satisfacción el sistema, es decir, el 80% manifestó estar de acuerdo con la usabilidad del sistema. Estas preguntas están relacionadas a temas sobre la simplicidad, comodidad, experticia y contenido. Estos resultados estarían vinculados a la necesidad de un tiempo mayor de capacitación sobre el manejo del sistema, para que los usuarios puedan conocer todas sus funcionalidades y se cuente con un periodo de práctica. Esta afirmación se puede respaldar por

lo indicado por Pérez et al. (2021), quien manifiesta que la capacitación es fundamental para que las tareas de mantenimiento sean eficaces. En este caso, mejorar la planificación de la etapa de capacitación para el manejo de esta herramienta informática es importante para lograr insertar correctamente este recurso en la gestión de mantenimiento del DMPQ de la empresa Proquímica. Se recomienda entonces brindar una capacitación del software, en términos de cómo incluir este recurso en las tareas de mantenimiento, indicando tanto como programar un cronograma y como mantener actualizada la información para garantizar la generación de la documentación completa de respaldo. Resulta importante considerar este hecho para evaluar la aplicación del sistema con una muestra mayor, además de evaluar la aceptación al cambio y el impacto en la carga de trabajo del personal.

Es necesario resaltar que para el buen manejo del sistema SGM-ELC es importante mantener actualizada la información e ingresarla correctamente, concordando con otros trabajos de desarrollo de sistemas para apoyar la gestión de mantenimiento (Ávila et al., 2022; Valenzuela 2023). Ávila et al. (2022) indica que el inventario de equipos y los reportes de mantenimientos son pilares fundamentales para el buen funcionamiento de un software para la gestión de mantenimiento y Valenzuela (2023) establece que la información de inventario, ficha técnica, registro de proveedores, registro de usuarios, programación de mantenimiento preventivo e historial de mantenimiento permite planificar adecuadamente el mantenimiento de equipos médicos.

Las respuestas recopiladas proporcionaron información valiosa orientada a la mejora del sistema SGM-ELC. De esta manera se puede proponer una segunda versión que brinde una herramienta para apoyar la gestión y la evaluación del mantenimiento, con acceso a información actualizada y a tiempo real, factores clave en la evaluación del mantenimiento de equipos médicos, como manifiesta Bahreini et al. (2019), Márquez (2023) y Zamzam et al. (2021a).

Como trabajo futuro, se propone mejorar el sistema con base a las recomendaciones de los participantes dando seguimiento al manejo del software para identificar los ajustes requeridos (Kumar et al., 2021). Posteriormente se planteará evaluar el impacto directo del uso del sistema SGM-ELC con todos los responsables de tareas de mantenimiento en la empresa Proquímica. Para esta última etapa se deberá considerar el adquirir datos a mediano y largo plazo, que permitan obtener métricas para conocer los porcentajes de la mejora en la gestión, planificación y ejecución del mantenimiento de los equipos, verificar la disminución de los gastos asociados a la reparación y la mejora en la calidad de atención a los clientes de la empresa. Información que debe ser evidenciada como parte de las mejoras de los servicios de las instituciones médicas y de los proveedores de equipos médicos (Badnjevic, 2023; Cuzco et al., 2019).

Conclusiones

La implementación de un sistema basado en una página web, una aplicación móvil y una base de datos permitió contar con una herramienta para la empresa Proquímica que puede ser incluida en la gestión del mantenimiento para acceder a información actualizada, automatizar la gestión de servicios, brindar seguimiento de órdenes de trabajo y registrar datos técnicos como apoyo en la gestión. El diseño se realizó con base en las necesidades de la empresa.

La evaluación realizada a partir de los resultados del cuestionario PSSUQ ha permitido comprender mejor la percepción y aceptación del sistema y la aplicación de esta propuesta. Se ha definido que es necesario mejorar la estrategia para capacitar al personal que usará el sistema. Se recomienda hacer énfasis en brindar acompañamiento al grupo humano que no se encuentra

familiarizado con el manejo de recursos informáticos para que esta herramienta de software pueda ser incluida de manera satisfactoria como recurso para la gestión de mantenimiento.

En el marco del proceso de mejora del DMPQ de la empresa Proquímica, se propone como trabajo futuro realizar la evaluación del sistema a largo plazo, con una muestra mayor para obtener estadísticas que verifiquen el impacto positivo de la aplicación del sistema SGM-ELC en la empresa.

Agradecimientos

Los autores expresan su agradecimiento al Departamento de Mantenimiento de la empresa Proquímica y a sus autoridades por su apoyo para la ejecución de este proyecto.

Referencias

- Ávila Cochancela, S. S., & Crespo Guillén, S. G. (2022). *Diseño plan de mantenimiento preventivo para equipos médicos del Hospital Santa Inés* (Bachelor's thesis).
- Badnjevic, A. (2023). Evidence-based maintenance of medical devices: Current shortage and pathway towards solution. *Technol. Health Care*, 31, 293-305.
- Bahreini, R., Doshmangir, L., & Imani, A. (2019). Influential factors on medical equipment maintenance management: In search of a framework. *Journal of Quality in Maintenance Engineering*, 25(1), 128-143.
- Cuzco, M. V., Villacrés-Parra, S., Gallegos-Londoño, C., & Negrete-Costales, H. (2019). Evaluación de la gestión del mantenimiento en hospitales del instituto ecuatoriano de seguridad social de la zona 3 del Ecuador. *Ingenius. Revista de Ciencia y Tecnología*, (22), 59-71.
- Hajesmaeel-Gohari, S., Khordastan, F., Fatehi, F., Samzadeh, H., & Bahaadinbeigy, K. (2022). The most used questionnaires for evaluating satisfaction, usability, acceptance, and quality outcomes of mobile health. *BMC Medical Informatics and Decision Making*, 22(1), 22. <https://doi.org/10.1186/s12911-022-01764-2>
- Iadanza, E., Gonnelli, V., Satta, F., & Gherardelli, M. (2019). Evidence-based medical equipment management: a convenient implementation. *Medical & biological engineering & computing*, 57, 2215-2230.
- Kumar, J. N., Pankaj, B., Kuldeep, S., Vibha, J., & Praveen, S. (2021). Assessment of computerized medical equipment management and maintenance System from health technology assessment (HTA) perspective: need and approach. *Health Technology Assessment in Action*.
- Márquez, F. P. G. (2023). *Operations Management and Management Science*.
- Pérez Hernández, Y., Vanega Osorio, R., Montero Sotolongo, F., Isaac, Z. J., & Caveda Medina, O. (2021). La capacitación del electromédico en gerencia para el mantenimiento a equipos médicos. *Educación Médica Superior*, 35(1).
- Shamayleh, A., Awad, M., & Farhat, J. (2020). IoT based predictive maintenance management of medical equipment. *Journal of medical systems*, 44, 1-12.
- Silva, J. P. V., Salas, C. J. L., Cordero, M. F. M., Forero, L. J. A., & Pereira, C. J. A. (2021, August). Development of predictor software to determine the periodicity of patient monitors preventive maintenance in the intensive care unit of a health care institution. In *2021 2nd Sustainable Cities Latin America Conference (SCLA)* (pp. 1-5). IEEE.
- Valenzuela Valenzuela, J. M. (2023). *Sistema de información para la gestión del mantenimiento del equipamiento biomédico del Hospital Regional del Cusco*.

Villa-Parra, A. C., Criollo, I., Valadão, C. A. A., Da Silva, L. F., Coelho, Y. L., Lampier, L., Rangel, L., Sharma, G., Delisle-Rodriguez, D., Calle-Siguencia, J., Urgiles-Ortiz, F., Díaz, C. A. R., Caldeira, E., Krishnan, S., & Bastos-Filho, T. (2022). Towards Multimodal Equipment to Help in the Diagnosis of COVID-19 Using Machine Learning Algorithms. *Sensors*, 22(12), 4341. <https://doi.org/10.3390/s22124341>

World Health Organization. (2012). *Introducción al programa de mantenimiento de equipos médicos*. Organización Mundial de la Salud.

Zamzam, A. H., Abdul Wahab, A. K., Azizan, M. M., Satapathy, S. C., Lai, K. W., & Hasikin, K. (2021a). A systematic review of medical equipment reliability assessment in improving the quality of healthcare services. *Frontiers in Public Health*, 9, 753951.

Zamzam, A. H., Al-Ani, A. K. I., Hasikin, K., & Wahab, A. K. A. (2021b, July). A Strategic Corrective Maintenance Prioritization Assessment for Medical Equipment. In *Kuala Lumpur International Conference on Biomedical Engineering* (pp. 133-141). Cham: Springer International Publishing.

Web app for the stimulation of auditory skills in children

Aplicación web para la estimulación de las habilidades auditivas en niños

Fabian Carvajal¹ <https://orcid.org/0000-0002-1303-5582>, Esteban Crespo-Martinez¹ <https://orcid.org/0000-0002-3061-9045>, Catalina Astudillo-Rodriguez¹ <https://orcid.org/0000-0001-8369-5300>,
Norma Reyes¹ <https://orcid.org/0000-0002-6723-1423>

¹Universidad del Azuay, Cuenca, Ecuador

fabianc@uazuay.edu.ec, ecrespo@uazuay.edu.ec,
cvastudillo@uazuay.edu.ec, nreyes@uazuay.edu.ec



Esta obra está bajo una licencia internacional
Creative Commons Atribución-NoComercial 4.0

Sent: 2023/06/20

Accepted: 2023/08/22

Published: 2023/10/15

Abstract

The existing literature identifies various situations related to hearing problems in children, such as difficulties in understanding oral language, delayed speech development, or auditory discrimination problems. In this work, we present a technological tool designed to stimulate listening skills in children from 5 to 7 years old, following the principles of Condemarin. This tool offers stimulation activities and provides immediate feedback to improve the ability to recognize, discriminate, and interpret auditory stimuli, which contributes to the development of communication skills in children. The tool we developed is a web application that is easy for therapists to install, learn, and configure. For its creation, we follow the prototype methodology, which includes the requirements, assembly, evaluation, corrections, documentation, and testing stages. We collaborate with a multidisciplinary team of experts in information technology, linguistic rehabilitation, education, and graphic design. Throughout the six phases of our process, we kept user needs in mind, resulting in a system evaluation that demonstrated high user satisfaction and good performance both in technical and interface terms. In addition, this application benefits children with language disorders and learning difficulties and children from 5 to 7 years old since it encourages the development of their communication skills.

Keywords: Auditory software, aural skills, prototype for listening skills, stimulation software, web app.

Summary: Introduction, Methods, Results of the instantiation of the methodology and Conclusions.

How to cite: Carvajal, F., Crespo-Martinez, E., Astudillo-Rodriguez, C., & Reyes, N. (2023). Web app for the stimulation of auditory skills in children. *Revista Tecnológica - Espol*, 35(2), 109-121. Recuperado a partir de <http://www.rte.espol.edu.ec/index.php/tecnologica/article/view/1022>

Resumen

En la literatura existente, se identifican varias situaciones relacionadas con problemas auditivos en niños, tales como dificultades en la comprensión del lenguaje oral, retrasos en el desarrollo del habla o problemas de discriminación auditiva. En este trabajo, presentamos una herramienta tecnológica diseñada para estimular las habilidades auditivas en niños de 5 a 7 años, siguiendo los principios de Condemarin. Esta herramienta ofrece actividades de estimulación y proporciona retroalimentación inmediata para mejorar la capacidad de reconocer, discriminar e interpretar estímulos auditivos, lo que contribuye al desarrollo de las habilidades de comunicación en los niños. La herramienta que desarrollamos es una aplicación web de fácil instalación, aprendizaje y configuración para los terapeutas. Para su creación, seguimos la metodología del Prototipo, que incluye las etapas de requerimientos, montaje, evaluación, correcciones, documentación y pruebas. Trabajamos en colaboración con un equipo multidisciplinario compuesto por expertos en tecnologías de la información, rehabilitación lingüística, educación y diseño gráfico. A lo largo de las seis fases de nuestro proceso, tuvimos en cuenta las necesidades de los usuarios, lo que resultó en una evaluación del sistema que demostró una alta satisfacción de los usuarios y un excelente rendimiento tanto en términos técnicos como en la interfaz. Además, esta aplicación beneficia tanto a niños con trastornos del lenguaje y dificultades de aprendizaje como a niños de 5 a 7 años, ya que fomenta el desarrollo de sus habilidades de comunicación.

Palabras clave: Aplicación web, habilidades auditivas, prototipo de habilidades auditivas, software de estimulación, software auditivo.

Introduction

The World Health Organization (WHO) has identified significant barriers to access to speech and language treatment in terms of cost, long-term participation, and difficulty in extending interventions beyond the context of telemedicine. Several studies have explored using systems to overcome these barriers through technology with untapped potential to address these issues, enabling them to access quality rehabilitation services through their necessary resources. Telemedicine uses information and communication technologies (ICT) to provide remote medical services, allowing medical consultations, diagnoses, treatments, and follow-ups to be performed digitally. This may include using medical devices connected to the Internet, video conferencing, email, messaging, and other communication tools (Gačnik et al., 2018; Morales-Acosta, 2020; Spitale et al., 2021).

These technologies aim to help people improve access to information and communication through specific devices and programs. Interventions for speech therapy therapeutic purposes using technological means, whether virtual platforms and/or software programs, have shown significant improvements in the linguistic, cognitive, social, emotional, and behavioral development of people in therapy, evidencing the increase in motivation and participation in different therapeutic activities, mainly playful, whether these are in teleconsultation or in virtual platforms or applications in electronic devices, which in turn has affected behavior, both personal and in the ability to analyze and detect behavioral patterns in other people (Buitrago et al., 2019; Gerosa et al., 2009; Spitale et al., 2021; Walden, 2021).

Speech-language pathology is a medical discipline dedicated to studying, diagnosing, and treating problems related to speech, language, hearing, and balance (Isaías & Campra, 2021).

Telemedicine, for its part, has allowed speech and language pathologists to reach patients who live in remote areas or who have difficulties traveling due to physical or cognitive

disabilities. To guarantee the quality and safety of the services provided, it is essential to regulate telemedicine and remote practice with the support of assistive communication technologies, such as screen readers, adapted keyboards, voice synthesis devices, and sign language translation software, which can help people with disabilities improve their access to information and communicate with others (Acosta et al., 2020).

Literature Review

The Internet has revolutionized many areas of modern life since it became a global means of communication. Information and Communication Technologies (ICT) offer the option of analyzing, organizing, synthesizing, and transmitting information and, therefore, can contribute to expanding and developing valuable clinical and educational tools for speech and language (Toki et al., 2012). In addition, it allows us to offer personalized and quality service (Lavoie et al., 2020).

"Simone Says" is computer software designed for children with ASD to acquire functional language through an interactive environment for language recovery that brings together research on speech recognition, natural language processing, and computer-assisted instruction, allowing concentration and primarily visual and auditory memory to work (Lehman, 1998).

The study of Glykas and Chytas (2004) describes the first web-based tool called Telelogos for diagnosis, treatment, and digital learning in speech and language therapy. It consists of three modules: public awareness, referral, and evaluation, which allow speech therapists and other professionals to find the optimal treatment for each patient. The tool was evaluated by twenty speech and language therapists and tested by potential users in Greece and the UK. The evaluation results were encouraging and demonstrated user satisfaction regarding efficiency and functionality. According to Griffiths et al. (2007), "high user satisfaction" means that users or customers of a product, service, or experience are very satisfied or pleased with it. High user satisfaction typically indicates that people find the product or service to be effective, enjoyable, and meeting their expectations or needs. It suggests a favorable response from the users, often seen as a desirable outcome for businesses and organizations aiming to provide quality products or services.

The authors, Pierrakeas et al. (2005), present the use of collaborative online environments (telemedicine) for speech therapy, in which the therapist can guide the group of participants through structured discussions, focusing on improving the initiation of speech. Conversation, taking turns, clarifying ideas, repairing breaks in conversation, and participating in role-plays of everyday communication situation. In addition, these environments employ easy-to-use recording, editing, and storage functions, so users can create professional recordings at their own pace in industry-standard formats using a standard computer with a camera and speakers. One benefit of these online collaboration environments is that they can support an unlimited number of co-presenters and subject-matter experts.

Toki et al. (2012) developed an online speech-language pathology expert system for diagnosing speech disorders and difficulties. The system can answer specific questions, give an automatic evaluation response, and generate a list of localized disorders. SLPs and other professionals can use this feedback to plan therapy or learning tasks.

There is a large amount of research regarding methods and technological resources implemented in speech therapy. The most commonly used technological resources are video calls, 2D and 3D resources, multimedia, voice recognition, and speech-to-text for patients to

obtain the necessary feedback for the correct development of the therapeutic session. Several investigations suggest that these tools are most accessible and convenient for everyone (Lo et al., 2017). In addition, they support patient-centered and self-management models of medical care (Madeira et al., 2017), significantly influencing the socio-emotional, motivational, and cognitive aspects (Buitrago et al., 2019).

In work carried out by Buitrago et al. (2019), a technological device is presented to support the therapeutic processes of children with learning problems to develop the following skills: memory, concentration, mathematical operations, laterality, and spatial orientation. The device consists of a mobile application for Android devices, three hardware interfaces, and a Lego mobile robot. The app has three activities (challenge, math, and maze). After the validation, it was possible to confirm an improvement in the children's abilities in the activities (memory, concentration, mathematical operations, laterality, and spatial orientation), as well as improvements in their self-esteem and general performance.

The main objective of the study by Lavoie et al. (2020) was to investigate the efficacy of a self-administered therapy for people with primary progressive aphasia (PPA) using a smart tablet to improve function word naming and assess generalization to a conversation task. The treatment was administered to five adults diagnosed with PPA four times a week for four consecutive weeks. As a result, a significant improvement was determined for the trained words in the five participants, and the gains were maintained two months after the treatment in four of them. In addition, evidence of generalization in conversation was found in three participants. Two aspects of the system stand out: i) a self-applicable type of therapy can be carried out where the person has control over their therapeutic process and their subsequent performance, and ii) it makes it possible to keep a detailed record of the evolution of the patient in different aspects of the therapy.

In the study carried out by Catania et al. (2022), a system was designed and developed that encourages children to describe themselves through questions and suggestions to create an avatar consistent with what is said, thought, and seen in reality, allowing the user to have a positive impact in different areas such as i) self-awareness, ii) self-knowledge, iii) communication skills, and iv) vocabulary of the mother tongue and foreign language.

Besides, Condemarin et al. (2016) identified five critical areas of auditory stimulation: i) Sound Recognition Stimulation, Activities to identify and recognize diverse sounds, strengthening hearing capacity; ii) Stimulation of Auditory Discrimination: Development of the ability to differentiate similar sounds, improving speech comprehension; iii) Sound Localization Stimulation: Helps determine the direction of sounds, crucial for spatial awareness; iv) Stimulation of Auditory Comprehension: Improves the interpretation and processing of auditory information, facilitating communication; and v) Stimulation of Auditory Memory: Reinforces the retention of auditory information in the short and long term.

Regarding this software, there is identified software with similar purposes available on the market. Among others: i) Lingokids is primarily focused on learning the English language and providing an engaging educational experience for children; ii) CogniFit is an application that offers auditory stimulation exercises and other cognitive activities to improve general auditory and cognitive skills. This app is available in multiple languages, including Spanish, and offers a free version with limited features and a paid version that provides complete access to all features and exercises. iii) Auditory Processing Studio is an application designed to help people with central auditory processing difficulties. It provides a variety of ear training exercises and

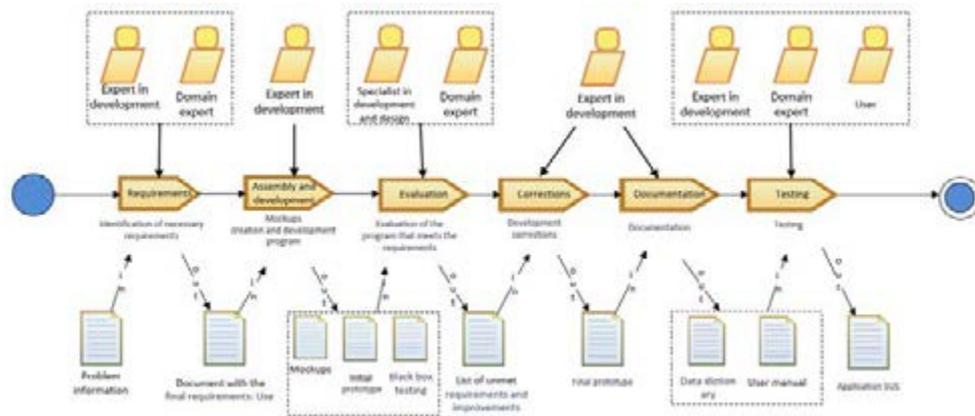
is available in multiple languages, including Spanish. This app offers a free trial version, but it also has a paid version with additional features and more advanced exercises.

Methods

The prototyping methodology allows the developer to create, refine, and test a working version. Despite the increasing focus on the detailed and controlled administration of the design process of a system, the prototyping methodology could offer data on the application's performance generated in the design phase (Armenta Benitez et al., 2018).

The prototype model of Armenta Benitez et al. (2018) was considered to develop the software prototype that supports the stimulation of auditory skills in children from 5 to 7 years old. Figure 1 (SPEM 2.0 (OMG, 2008)) visualizes the methodology applied in the construction of the prototype. This methodology consisted of six phases: i) Requirements, ii) Assembly and development, iii) Evaluation, iv) Documentation, and v) Testing.

Figure 1
Stages of the prototyping methodology



Results of the instantiation of the methodology

Requirements Phase

The Requirements phase is developed in a unidirectional way between three activities adapted to the domain of this study (Amado-Salvatierra et al., 2015): i) definition of the problem, which includes i) defining the auditory stimulation program (guidelines proposed by Condemarin) and formulate the objectives to be achieved; ii) identification of requirements: identify the activities to be carried out, define the rules to be followed to guarantee accessibility/usability, and select or create graphic and multimedia resources, for which the use case diagram was defined; and iii) use cases: where the requirements for the implementation of the prototype are defined. As a result of this phase, the use cases were defined based on the use case diagram and stimulation modules, as shown in Tables 1 and 2 and Figure 2.

Table 1
Stimulation modules

CODE	STIMULATION MODULES	DESCRIPTION
ACT1	Module 1 Activities	Auditory memory
ACT2	Module 2 Activities	Auditory Awareness

CODE	STIMULATION MODULES	DESCRIPTION
ACT3	Module 3 Activities	Auditive discrimination
ACT4	Module 4 Activities	Initial sounds
ACT5	Module 5 Activities	final sounds

Figure 2
Use Case Diagram

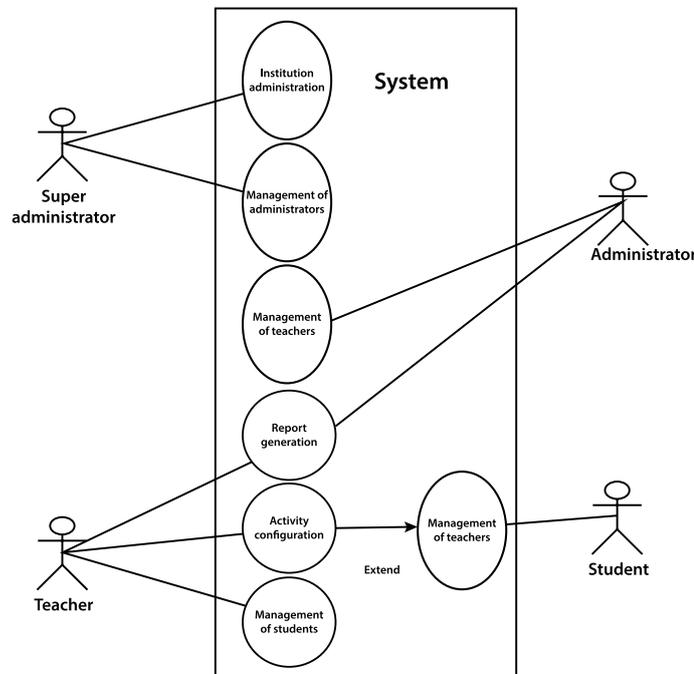


Table 2
Use Cases

Code	Case of use	Actor	Description	Priority
CU1	Administration of Institutions	Super Admin	It allows one to take control of the different institutions.	Mandatory
CU2	Management administrators	Super Admin	It allows one to take control of the different administrators who will take control of the platform.	Mandatory
CU3	Administration of teachers	Administrator	Allows one to take control of the different teachers who will take control of the stimulation activities.	Mandatory
CU4	Generation of reports	Administrator, Teacher	It allows one to take control of the reports, with the results for the students when carrying out the activities. Stimulation.	Necessary
CU5	Configuration of stimulation activities	Teacher	It allows one to configure the necessary stimulation activities for students (ACT 1 - ACT 7).	Mandatory
CU6	Carry out stimulation activities	Students	It allows one to carry out the stimulation activities generated by the teacher (ACT 1 - ACT7).	Necessary
CU7	Student Administration	Teacher	It allows one to keep track of the different students who carry out the activities. Stimulation.	Mandatory

Assembly and development phase

Based on use cases, the second phase was developed unidirectionally between two activities adapted to the domain of this study: i) Assembly and ii) Development (Amado-Salvatierra et al., 2015). In this way, a generic plan was created for how the software will be

built, which includes choosing the optimal instructional method or methods to create objects or materials that help support the therapeutic processes of children with learning disabilities (Buitrago et al., 2019).

Assembly Activity: As a result of this activity, the interfaces (mockups) were designed for the activities ACT1 to ACT7.

Development Activity: The architecture used to implement the web prototype is based on the Model View Controller (MVC) (Rojas et al., 2019). The implementation of this model was based on the platform (Yii, PHP, MARIADB, speech recognition, and the Visual Studio Code IDE) to develop a web application compatible with mobile devices, tablets, and computers. As a result of this activity, the first version of the software prototype was obtained with the Use Cases CU1 to CU7 implemented. During the internal tests of the software among the developers, performance evaluations were made for each implemented item. As an output, the partial solution software prototype was obtained.

Evaluation Phase

The evaluation phase is carried out in a unidirectional way between three activities (Amado-Salvatierra et al., 2015) adapted to the core of this study: i) Planning the evaluation; ii) Gathering information; iii) Analysis of the information obtained. These activities determine if the software prototype achieves the planned objectives and what changes are necessary.

For this phase, it was considered necessary to evaluate the requirements fulfilled and the system's functionality through the type of software tests known as "black boxes". The input was the first version of the prototype obtained in the previous stage. The checklists for evaluating fulfilled requirements and functionality were based on the model (Aristegui, 2010). Tables 3 and 4 show the compilation of the evaluation carried out by the expert in the field of speech therapy.

Table 3
Requirements Assessment Checklist

Requirements	Met the requirement		Comments observations
	YES	NO	
The system will allow the creation of different stimulation activities the teacher generates.	X		Works correctly
The system will allow one to view the options of each stimulation activity randomly.	X		Works correctly
The system will allow one to reproduce the texts of the answers in audio form.	X		Works correctly
The system will allow the playback of the audio files uploaded by the teacher to stimulate the activity.	X		Works correctly
The system will allow, through voice recognition, to select the response of the stimulation activity in the form of an image.	X		Works correctly
The system will allow, through voice recognition, to say the answer.	X		Works correctly
The system will allow visualizing feedback between each stimulation activity.	X		Works correctly

Note: Table 3 provides eight successful sample request requests that can be performed on the prototype. Therefore, the success rate for evaluating fulfilled system requirements is 100% (Sasmito & Nishom, 2020).

Table 4
Functionality Assessment Checklist

Action	Scenery	Met the requirement		Comments observations
		YES	NO	
Does the system start without a problem?	The student enters his username, password, and institution	X		Works correctly
	The student presses continue and is redirected correctly.	X		Works correctly
	The student accesses the system.	X		Works correctly
Does the system display the correct information?	When accessing the system, the correct student information is displayed.	X		Works correctly
Does the system allow stimulation activities to be carried out correctly?	Auditory awareness module.	X		Works correctly
	Auditory memory module.	X		Works correctly
	Auditory discrimination module.	X		Works correctly
	Initial sounds module.	X		Works correctly
	Final sounds module.	X		Works correctly
Does voice recognition work correctly?	By pressing the voice recognition button, an answer can be selected .	X		Works correctly
In stimulation activities, does the system reproduce the orders?	Pressing the play button plays the command.	X		Works correctly
Does the system show feedback on the stimulation activities carried out?	Upon completing the stimulation activities, it displays a message with "Well done."	X		Works correctly
	At the end of the stimulation activity, it shows the general result.	X		Works correctly
Does the system display the information according to roles?	Super Admin	X		Works correctly
	Administrator	X		Works correctly
	Teacher	X		Works correctly
	Student	X		Works correctly

Note: Table 6 provides seven successful sample actions that can be executed on the software prototype. Therefore, the success rate for the functionality of the system is 100% (Sasmito & Nishom, 2020).

During the evaluation, five observations for improvements are listed in Table 6.

Table 5
List of improvements

Improvement number	Improvements to perform	Roles affected
1	Correction of the list in user report module: Display a list of users in the user report.	Super Admin Administrator Teacher
2	Filter correction in the general report module: Implement a search engine in the general report to filter the report by a teacher.	Super Admin Administrator Teacher
3	Correction of the list in the general report module: Implement a table in the general report to view the teachers.	Super Admin Administrator Teacher

Improvement number	Improvements to perform	Roles affected
4	Correction of information in the general report module: Eliminate unnecessary data in the general report.	Super Admin Administrator Teacher
5	Correction of behavior for teacher role in user module: Disable the option to select the institution and manager of the user when creating the student.	Teacher

Corrections phase

Based on the initial prototype and the list of improvements as inputs, the correction phase is carried out in a unidirectional way in four activities adapted to this study (Amado-Salvatierra et al., 2015): (1) daily meetings; (2) reuse/adaptation of pre-existing material; (3) Development/Modification; and iv) Software error testing.

As a result of this phase, the Final Prototype was obtained, shown in Figure 3, and accessible at the following link: <https://github.com/luchV/PrototipoSoftware.git>.

Figure 3
Final prototype



Documentation Phase

The documentation phase is carried out in a unidirectional way in two activities (Amado-Salvatierra et al., 2015) adapted to the project: i) Preparation of the Data Dictionary and ii) Preparation of the User Manual. In this phase, the documentation that will allow us to understand the operation of the prototype is generated. As an output of this phase, the data dictionary and user manual were obtained (<https://n9.cl/manual-de-usuario>).

Implementation phase

The implementation phase is executed in a unidirectional way between three activities (Amado-Salvatierra et al., 2015): i) Perform software usability tests; ii) Organization technical support; and iii) Retrospective. This way, the prototype created in the Assembly and Development phase was tested in the field (Cheung, 2016). As a result of this phase, the final Prototype Usability report was obtained.

A survey was applied that was based on the System Usability Scale (SUS) (Bangor, 2009). Google Forms was used to extract the information using a 5-item Likert scale. In this case, the participants were 36 students from the Universidad del Azuay in the Educational Psychology career who carried out a representative sample of tasks for the product (usability tests) and then completed the survey before any discussion with the moderator.

The statistical analysis process was based on the work of (Cedillo Orellana, 2017) . The collected results validated the questionnaire through Cronbach's alpha, obtaining a value of 0.81. Since Cronbach's alpha was more significant than 0.7, it suggests that the questionnaire is valid and consistent in measuring the desired construct.

The usability measurement scale (SUS) was used to obtain the descriptive statistics through the Excel 2019 software. The result was 85.6, so the prototype corresponds to the " excellent " category with high usability. (Bangor, 2009).

This application is a web platform that will be hosted on the servers of the Universidad del Azuay. It is important to note that access to the application is free for public and private institutions in Cuenca. In this way, we ensure the application is accessible and beneficial to a broad spectrum of users in the city's educational community.

Conclusions

To respond to the research question, which focuses on the skills considered in developing software for auditory stimulation, observing that, according to the literary review, a significant percentage of the articles (42%) focus on activities to stimulate auditory skills, such as detection, localization, discrimination, and auditive identification. Another 42% is related to activities that stimulate linguistic and extra-linguistic skills, such as attention, concentration, memory, behavior, and interpersonal relationships. In addition, 24% of the articles address the control and management of frustration and the teaching of vocal acoustics, among other aspects. Based on the literature review, these findings align with the listening skills proposed by Condemarin and provide valuable information for the development of auditory stimulation software.

The web application prototype to stimulate listening skills is beneficial for children with speech disorders and learning difficulties and typically developing children between 5 and 7 years of age. The application offers stimulation activities and immediate feedback to improve the ability to recognize, discriminate, and interpret auditory stimuli, which can help develop communication in children. In addition, this prototype has great potential to complement traditional therapies and expand therapists' ability to meet their patients' needs.

Unlike other auditory stimulation applications, the developed prototype stands out for its voice recognition feature, which provides notable benefits: i) Practice Pronunciation: Speech recognition helps users perfect their pronunciation of words and sounds, providing accurate and effective feedback; and ii) Interactive instructions: Especially in apps for children, interactive verbal instructions make activities more dynamic and personalized, allowing users to give voice commands and receive interactive responses.

The system evaluation results showed high user satisfaction and excellent efficiency in technical and interface terms. Users found the system extremely useful due to its ability to customize stimulation activities and quickly generate reports. In general, the prototype is believed

to improve the quality of speech and language therapy services thanks to the implementation of advanced technologies such as voice recognition.

This tool is intended to significantly improve listening comprehension in high school students, a crucial skill for linguistic learning and communication. It is adapting the activities proposed by Condemarín in areas such as auditory awareness (recognizing natural and human-produced sounds), auditory memory (repeating words and associating them with images, verifying animal sounds), auditory discrimination (differentiating animal sounds and environment), and the recognition of initial, incorrect, and final sounds in words. Importantly, consistently fostering these listening skills from an early age is essential for language development, knowledge acquisition, and academic success.

Limitations and future works

In this first stage, the application was used only by professors acting as testers. In the next stage of this research, we must include kids to evaluate their reactions to the solution.

Regarding security, the code does not incorporate significant security measures like data input validation or attack prevention. It is essential to implement adequate security measures, mainly if the code handles sensitive information or interacts with users. For example, it is recommended to use the functions and methods provided by the Yii software framework to manage sessions, as they ensure the security and integrity of the application. Furthermore, validating and filtering all data inputs is essential to preventing vulnerabilities such as SQL injection or cross-site scripting (XSS) attacks.

The implementation of future improvements in the auditory stimulation web application includes:

1. **Advanced Hearing Evaluation:** It seeks to incorporate more advanced hearing evaluations that allow users to measure their listening skills more accurately, including their ability to comprehend and answer oral questions.
2. **Improvement in Voice Control:** It is intended to improve voice control functionality, mainly aimed at users with hearing disabilities, to further facilitate navigation and interaction through more precise and varied voice commands.
3. **Intelligent Adaptation:** The application will focus on developing a more intelligent adaptation to users' verbal responses. This will allow the app to adjust activities and difficulty levels more precisely, providing a more personalized and practical learning experience based on user performance.

These enhancements are intended to enrich the user experience and maximize the benefits of auditory stimulation through technology and voice interaction.

References

- Acosta, G., Cajiao, M., & Arias, A. (2020). Post-agreement and Deaf Diversity: He narrated the conflict in Colombian Sign Language (LSC). Inclusion of cultural and social diversity in education. *Center for Latin American Studies on Inclusive Education (CELEI)*.
- Amado-Salvatierra, HR, Quan Lainfiesta, LR, & Hernández Rizzardini, R. (2015). Methodological guide for the creation of accessible virtual curricular developments. *Angewandte Chemie International Edition*, 6 (11), 951-952.

- Aristegui, JL (2010). Test Cases in Software Testing. *Lampsakos*, 3, 27. <https://doi.org/10.21501/21454086.785>
- Armenta Benitez, LÓPEZ, Samue, RODRIGUEZ-ESPINOZA, I., & MEDINA-MUÑOZ, L. (2018). Application of the Prototyping Model: Case Study RedbotGamesShop Software. 2 (5 8-13).
- Bangor, A. (2009). Determining What Individual HIS Scores Mean: Adding an Adjective Rating Scale. 4 (3).
- Buitrago, JA, Andrade, CC, & Bravo, EC (2019). A new technological therapy device for children with learning disorders. *Proceedings of the 5th Workshop on ICTs for Improving Patients Rehabilitation Research Techniques*, 1-4. <https://doi.org/10.1145/3364138.3364140>
- Catania, F., Spitale, M., & Garzotto, F. (2022). Conversational Agents in Therapeutic Interventions for Neurodevelopmental Disorders: A Survey. *ACM Computer. Surv.* <https://doi.org/10.1145/3564269>
- Cedillo Orellana, IP (2017). *Monitoring the quality of cloud services through runtime models [Universitat Politècnica de València]*. <https://doi.org/10.4995/Thesis/10251/78032>
- Cheung, L. (2016). *Using the ADDIE Model of Instructional Design to Teach Chest Radiograph Interpretation. Journal of Biomedical Education*, 2016, 1-6. <https://doi.org/10.1155/2016/9502572>
- Condemarin, M., Gorostegui, M., Chadwick, M., & Milicic, N. (2016) . *Madurez Escolar*. Editorial Ediciones UC. Chile.
- Gačnik, M., Starčič, AI, Zaletelj, J., & Zajc, M. (2018). User-centred app design for speech sound disorders interventions with tablet computers. *Universal Access in the Information Society*, 17 (4), 821-832. <https://doi.org/10.1007/s10209-017-0545-9>
- Gerosa, M., Giuliani, D., Narayanan, S., & Potamianos, A. (2009). A review of ASR technologies for children's speech. *Proceedings of the 2nd Workshop on Child, Computer and Interaction - WOC CI '09*, 1-8. <https://doi.org/10.1145/1640377.1640384>
- Glykas, M., & Chytas, P. (2004). Technology assisted speech and language therapy. *International Journal of Medical Informatics*, 73 (6), 529-541. <https://doi.org/10.1016/j.ijmedinf.2004.03.005>
- Griffiths, J., Johnson, F. & Hartley, R. (2007). User satisfaction as a measure of system performance. *Journal of Librarianship and Information Science*. <https://doi.org/10.1177/0961000607080417>
- Isaías, A., & Campra, M. (2021). *Phonoaudiology: Possible Interventions and Practices*.
- Lavoie, M., Bier, N., Laforce, R., & Macoir, J. (2020). Improvement in functional vocabulary and generalization to conversation following a self-administered treatment using a smart tablet in primary progressive aphasia. *Neuropsychological Rehabilitation*, 30 (7), 1224-1254. <https://doi.org/10.1080/09602011.2019.1570943>
- Lehman, J.F. (1998). Toward the Use of Speech and Natural Language Technology in Intervention for a Language-disordered Population. 8.
- Lo, H.-C., Tang, S.-T., Wei, W.-L., & Chuang, C.-C. (2017). Design and Usability Evaluation of Speech Rehabilitation APP Interface for Patients with Parkinson's Disease. In M. Antona & C. Stephanidis (Eds.), *Universal Access in Human-Computer Interaction. Human and Technological Environments* (Vol. 10279, pp. 65-75). Springer International Publishing. https://doi.org/10.1007/978-3-319-58700-4_6
- Madeira, RN, Mestre, V., & Ferreira, T. (2017). Phonological Disorders in Children? Design and user experience evaluation of a mobile serious game approach. *Proceeds Computer Science*, 113, 416-421. <https://doi.org/10.1016/j.procs.2017.08.359>
- Morales-Acosta, G. (2020). Human Communication: Construction of an Intercultural Phonoaudiology in the Dimensions of Social and Cultural Interaction. *Inclusions Magazine*, 264-281.
- Pierrakeas, C., Georgopoulos, VC, & Malandraki, GA (2005). Online Collaboration Environments in Telemedicine Applications of Speech Therapy. *2005 IEEE Engineering in Medicine and Biology 27th Annual Conference*, 2183-2186. <https://doi.org/10.1109/IEMBS.2005.1616895>

- Sasmito, GW, & Nishom, M. (2020). Testing the Population Administration Website Application Using the Black Box Testing Boundary Value Analysis Method. *2020 IEEE Conference on Open Systems (ICOS)*, 48-52. <https://doi.org/10.1109/ICOS50156.2020.9293645>
- Spitale, M., Silleresi, S., Leonardi, G., Arosio, F., Giustolisi, B., Guasti, MT, & Garzotto, F. (2021). Design Patterns of Technology-based Therapeutic Activities for Children with Language Impairments: A Psycholinguistic-Driven Approach. *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*, 1-7. <https://doi.org/10.1145/3411763.3451775>
- Toki, EI, Pange, J., & Mikropoulos, TA (2012). An Online Expert System for Diagnostic Assessment Procedures on Young Children's Oral Speech and Language. *Proceeds Computer Science*, 14, 428-437. <https://doi.org/10.1016/j.procs.2012.10.049>
- Walden, PR (2021). Learning Acoustic Assessment of Voice: An Action Research-Based SoTL Project. *Journal of Voice*, S0892199721004057. <https://doi.org/10.1016/j.jvoice.2021.11.015>

Modelo de Visualización de Datos de Juegos Serios de Atención y Memoria Cognitiva

Data Visualization Model of Serious Games of Cognitive Attention and Memory

Juan-Sebastián Toledo^{1,2} <https://orcid.org/0000-0001-5120-9486>, Juan-Fernando Lima² <https://orcid.org/0000-0003-3500-3968>, María-Inés Acosta-Urigüen^{1,2} <https://orcid.org/0000-0003-4865-2983>,
Marcos Orellana² <https://orcid.org/0000-0002-3671-9362>

¹*Escuela de Ciencias de la Computación, Universidad del Azuay, Cuenca, Ecuador*
sebastiantoledo@es.uazuay.edu.ec

²*Laboratorio de Investigación y Desarrollo en Informática (LIDI), Universidad del Azuay, Cuenca, Ecuador, Cuenca, Ecuador*
flima@uazuay.edu.ec, macosta@uazuay.edu.ec,
marore@uazuay.edu.ec



Esta obra está bajo una licencia internacional Creative Commons Atribución-NoComercial 4.0

Enviado: 2023/07/15

Aceptado: 2023/08/22

Publicado: 2023/10/15

Resumen

En el campo de la psicología, los juegos serios se han transformado en herramientas digitales que permiten la aplicación de test psicológicos, el entrenamiento de competencias, y la detección de trastornos o patologías. Si bien los sistemas interactivos generan grandes cantidades de datos que pueden ser almacenados, surge la necesidad de identificar patrones de juego que permitan al especialista tomar decisiones basadas en datos. En este contexto, los modelos de visualización se han convertido en una herramienta moderna y precisa para solventar estas representaciones. El objetivo del presente trabajo es crear un modelo de visualización aplicado a datos extraídos de un juego serio orientado al entrenamiento de atención y memoria. Para ello, se propuso una metodología que permitió el desarrollo de un entorno unificado de análisis visual compuesto por tres tableros interactivos. Finalmente, el modelo fue evaluado a través del modelo de aceptación tecnológica, demostrando una fiabilidad sobresaliente.

Sumario: Introducción, Revisión de literatura, Método y Materiales, Resultados y Discusión y Conclusiones.

Como citar: Toledo, J. S., Lima, J. F., Acosta-Urigüen, M. I., & Orellana, M. (2023). Modelo de Visualización de Datos de Juegos Serios de Atención y Memoria Cognitiva. *Revista Tecnológica - Espol*, 35(2), 122-142. Recuperado a partir de <http://www.rte.espol.edu.ec/index.php/tecnologica/article/view/1056>

Palabras clave: datos cognitivos, juegos serios, modelo de visualización, comportamiento de jugadores, mapeo visual.

Abstract

In the field of psychology, serious games have been transformed into digital tools that allow the application of psychological tests, skills training, and the detection of disorders or pathologies. Although interactive systems generate large amounts of data that can be stored, the need arises to identify game patterns that allow the specialist to make data-driven decisions. In this context, visualization models have become a modern and precise tool to solve these representations. This work aims to create a visualization model applied to data extracted from a serious game aimed at training attention and memory. For this purpose, a methodology was proposed that allowed the development of a unified visual analysis environment composed of three interactive dashboards. Finally, the model was evaluated through the technology acceptance model, which showed reliability equivalent to outstanding.

Keywords: cognitive data, serious games, visualization model, players behavior, visual mapping.

Introducción

En los campos de la psicología y la medicina, los juegos serios se han transformado en herramientas digitales que permiten la aplicación de test psicológicos, el entrenamiento de competencias y la detección de trastornos o patologías, desempeñando un papel fundamental para la salud mental y proporcionando al especialista una efectiva herramienta de soporte a los métodos de terapia tradicionales (Abd-Alrazaq et al., 2022; Manera et al., 2017; Mezrar & Bendella, 2022). Al igual que cualquier otro sistema altamente interactivo, los juegos serios generan grandes cantidades de datos, que reflejan directamente las acciones y decisiones del jugador (Dörner et al., 2016). Las técnicas de minería de datos (MD), que hoy en día son muy comunes en diversos campos como la educación, medicina y finanzas (Kumar & Bhardwaj, 2011), pueden aplicarse a la gran cantidad de información que se deriva de la interacción de los usuarios con los juegos serios (Alonso-Fernández et al., 2019), de esta forma es posible identificar patrones de juego que permitan al especialista tomar decisiones basadas en datos (Provost & Fawcett, 2013).

Aún en la actualidad, es común que en el ámbito cotidiano los especialistas encargados de controlar y llevar a cabo las sesiones de entrenamiento no posean información detallada acerca del estado del juego, el progreso o las implicaciones que conlleva a las decisiones del jugador; y en consecuencia se trate al sistema como una caja negra¹ (Alonso-Fernandez et al., 2021). A pesar de los avances tecnológicos y las nuevas herramientas que permiten interactuar con la información recolectada de juegos serios, los datos generados no reciben un tratamiento adecuado y la falta de comunicación de hallazgos resulta en la pérdida de información relevante y una incorrecta gestión del conocimiento (Loh et al., 2015). En este contexto surge la necesidad de aplicar técnicas de MD que permitan registrar, procesar y analizar la información, con el propósito de desarrollar un modelo de visualización adecuado para juegos serios en el contexto de atención y memoria. El análisis de datos puede revelar creencias, comportamientos y estrategias de resolución de problemas del jugador, mientras que la visualización permite transformar la información en gráficos apropiados, para facilitar la interpretación de resultados a las partes interesadas (G. Wallner & Kriglstein, 2013).

¹ Sistema que produce resultados útiles sin revelar información acerca de su funcionamiento interno.

El presente trabajo se enfoca en aplicar técnicas de MD en un conjunto de datos provenientes de un juego serio orientado al entrenamiento de atención y memoria, con la finalidad de desarrollar un modelo de visualización interactivo, que permita al especialista de la salud explorar, analizar y evaluar a los participantes a través de sus características propias (atributos demográficos, socioeconómicos y conductuales) y aquellas variables que denotan el rendimiento durante cada sesión (puntaje y tiempo de duración).

Revisión de literatura

Durante la última década los desarrolladores de juegos serios han comenzado a recolectar grandes volúmenes de datos durante las sesiones de juego, con el fin de comprender el comportamiento del jugador y permitir la toma de decisiones acertada, lo cual conlleva al surgimiento de la analítica de datos en el contexto de los juegos serios (G. Wallner & Kriglstein, 2013), que según Alonso-Fernandez et al. (2021) se define como el análisis de datos y la extracción de información, y la analítica visual de juegos para apoyar a la interpretación de datos complejos (Günter Wallner et al., 2018). Una investigación comparable a este trabajo ha sido descrita por Minović et al. (2015). Aquí los investigadores presentan un modelo para la visualización del aprendizaje en juegos serios, que permite al profesor obtener una retroalimentación en tiempo real del desempeño del estudiante, con la finalidad de facilitar la toma de decisiones en áreas de gran importancia como el nivel de dificultad del juego o el camino de aprendizaje apropiado para el estudiante. Por otro lado, la herramienta permite al estudiante conocer su progreso en el juego. La técnica de visualización utilizada consiste en una modificación del diagrama de vista circular propuesto por (D. A. Keim et al., 2008).

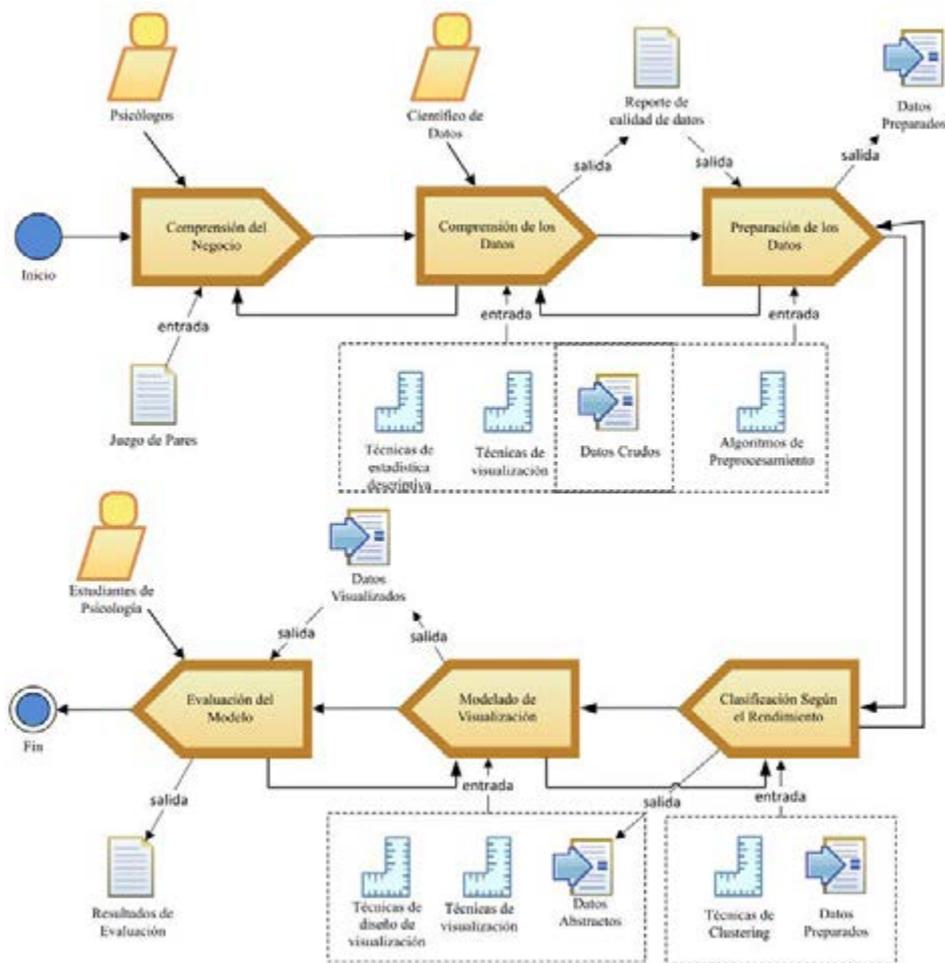
De Troyer et al. (2016) realizaron una investigación, que consistió en el desarrollo de tres técnicas de visualización que permiten informar acerca del estado del jugador y mejorar la comprensión de los resultados, en un juego serio orientado al ciberacoso en redes sociales. Cada una de las técnicas presentadas propone un enfoque específico y detallado en áreas de interés. La primera visualización se denomina *time-oriented-visualization*, y se enfoca en detallar la frecuencia relacionada a las interacciones del jugador con cada elemento del juego. La segunda, *character-oriented visualization*, se enfoca en el jugador y provee una visión general de las relaciones entre personajes en el juego. La tercera visualización, *interaction-oriented visualization*, resalta las interacciones entre jugadores. Como resultado, los investigadores demostraron que las visualizaciones ayudan a los participantes a comprender mejor el resultado del juego. La visualización que obtuvo la mejor puntuación en la evaluación fue *interaction-oriented visualization*, que se caracteriza por proporcionar un entorno interactivo para la exploración de las interacciones entre los jugadores, a través de una representación de nodos (jugadores) y aristas (interacciones entre jugadores).

Las técnicas de visualización utilizadas han demostrado alta efectividad para representar los datos según su propósito, entre las disponibles destacan: gráficos de pastel, gráficos de barras, gráficos de líneas y mapas de calor y gráficos de burbujas. Si bien existen investigaciones que utilizan técnicas de analítica de datos en juegos serios con diversos propósitos que abarcan desde la evaluación del aprendizaje hasta la validación del diseño del juego, no existen estudios que implementen un modelo de visualización adecuado para la representación de datos extraídos de un juego serio orientado al entrenamiento cognitivo, la atención y la memoria.

Método y Materiales

El presente trabajo se fundamenta en el estudio de la ciencia de datos y la visualización de información con la finalidad de desarrollar un modelo de visualización que permita identificar patrones entre las características propias de cada jugador (demográficas, socioeconómicas y conductuales), contra aquellas variables que denotan el rendimiento cognitivo (puntaje y tiempo) en el juego serio de pares. Para ello, se ha desarrollado una metodología basada en el método *Cross Industry Standard Process for Data Mining* (CRISP-DM) (Wirth & Hipp, 2000) y el *Modelo Unificado de Visualización* (MUV) de (Martig et al., 2003). Todo el proceso se ve reflejado en la Figura 1.

Figura 1
Metodología del macroproceso para el tratamiento, análisis y visualización de datos provenientes de juegos serios



El Juego de Pares es un juego en línea desarrollado por psicólogos y estudiantes de la Universidad “Nombre”, con el objetivo de evaluar y reforzar habilidades asociadas al área cognitiva (atención y memoria). La Figura 2 muestra parte del entorno de ejecución. El juego consiste en encontrar el par de cada imagen, para esto las imágenes se muestran un determinado tiempo y luego se ocultan, con el fin de que el jugador recuerde la posición inicial. Para completar el juego se deben superar cuatro niveles, cada uno con dificultad incremental (dos pares más por cada nivel). El rendimiento del jugador se determina por el tiempo empleado para completar cada nivel y el puntaje final (penalizado por cada error) con un máximo de 2000 puntos. Se puede acceder al juego mediante un navegador web, a través de la URL: <https://jserionew-8e818.web.app>.

Figura 2
Juego de Pares

Descripción del conjunto de datos

Los datos recolectados corresponden a sesiones registradas desde mayo del 2022 hasta septiembre del 2022 y se dividen en dos categorías. 1) Aquellos adquiridos mediante un único registro de usuario, en el cual se solicitaron datos de tipo demográficos, socioeconómicos y conductuales. 2) Datos generados durante la ejecución del juego, aquellos que denotan directamente la interacción del usuario con el sistema, como el tiempo de juego y el puntaje. En el estudio participaron un total de 248 personas, de las cuales 129 pertenecen al género femenino y 119 al género masculino. En la Tabla 1 se presentan las variables recolectadas.

Tabla 1
Descripción de las variables recolectadas del juego de pares

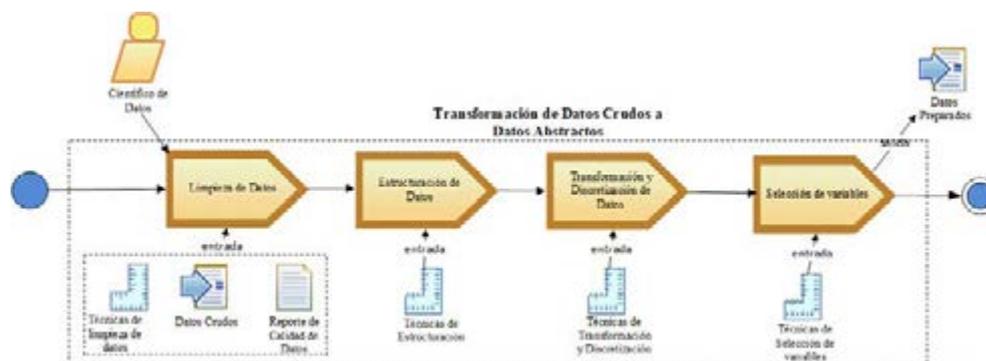
Atributos	Tipo de Dato
Cédula	Nominal
Nombre	Nominal
Ciudad	Nominal
Género	Nominal
Peso	Numérico
Altura	Numérico
Discapacidad	Booleano
Tipo Discapacidad	Nominal
Antecedentes neurodegenerativos	Booleano
COVID-19	Booleano
Fecha de nacimiento	Fecha
Nivel de Instrucción	Ordinal
Tipo de colegio	Nominal
Uso del computador	Numérico
Actividad física	Numérico
Ingresos económicos	Ordinal

Atributos	Tipo de Dato
Fecha de juego	Fecha
Tiempo nivel 1	Fecha
Tiempo nivel 2	Fecha
Tiempo nivel 3	Fecha
Tiempo nivel 4	Fecha
Tiempo total	Fecha
Puntos	Numérico

Preparación de los datos

La aplicación de técnicas de preprocesamiento previo al modelado puede incrementar la eficiencia del proceso y mejorar la calidad de los patrones encontrados. Es común que los datos recolectados no se encuentren en un estado óptimo, por lo tanto, deben ser tratados considerando las observaciones detectadas en la fase previa (Han, 2012). En la Figura 3, se presenta el proceso de preparación de los datos, que a su vez corresponde con la primera transformación del MUV.

Figura 3
Modelado de la preparación de los datos



La Transformación de Datos Crudos a Datos Abstractos (DC – DA) es la responsable de procesar los datos provenientes del dominio de aplicación (datos crudos) y llevarlos a un formato manejable por el sistema, y como resultado se obtiene un conjunto de datos potencialmente visualizables denominados datos abstractos (Martig et al., 2003). Como parte de los datos abstractos podemos tener un subconjunto de Datos Crudos a los cuales se les puede haber mejorado de alguna forma (Martig et al., 2003). En esta fase, se procesaron los datos abstractos mediante tareas de limpieza, estructuración, discretización y selección de variables. Como resultado, se obtuvo un conjunto de datos preparados que servirán como punto de partida para la continuación del proceso de transformación DC – DA, misma que finalizará en la siguiente fase (clasificación según el rendimiento).

Limpieza de datos

La limpieza de datos se puede aplicar para eliminar el ruido, tratar los valores atípicos y corregir las inconsistencias en los datos (Kotu & Deshpande, 2019). En este contexto, se realizó una optimización de los datos, mediante la corrección de: 1) unidades de medida y tiempo erróneas correspondientes a los atributos: “peso”, “altura”, “uso del computador” y “actividad física”. 2) Expresiones que contienen caracteres (P.ej. “aprox”, “h”, “cm”), en campos numéricos. 3) Corrección de “fecha de juego”, para ello se restó seis horas de cada campo registrado por el servidor. 4) Eliminación de registros con inconsistencias en la fecha de nacimiento, únicamente aquellos que ingresaron la fecha actual al momento del registro. Como resultado del proceso de

limpieza de datos se obtuvieron un total de 471 registros.

Discretización de datos

En este paso del preprocesamiento, los datos se transforman o consolidan para que el proceso de minería resultante sea más eficiente y los patrones encontrados se puedan interpretar con mayor facilidad (Han, 2012). Como parte del proceso se discretizó la edad de los jugadores según la clasificación de edades sugerida por la psicología del desarrollo (Wertsch & Tulviste, 1992), y los rangos asignados a cada grupo etario se pueden observar en la Tabla 2. Finalmente, se realizó la discretización del índice de masa corporal de acuerdo a la escala sugerida por el Centers for Disease Control and Prevention (Division of Nutrition, Physical Activity, and Obesity, 2022), los rangos empleados se pueden apreciar en la Tabla 3.

Tabla 2
Clasificación por grupos etarios

Edad (años)	Grupo etario
2 - 11	Niñez
12 - 19	Adolescencia
20 - 35	Juventud
36 - 60	Madurez
61 o más	Vejez

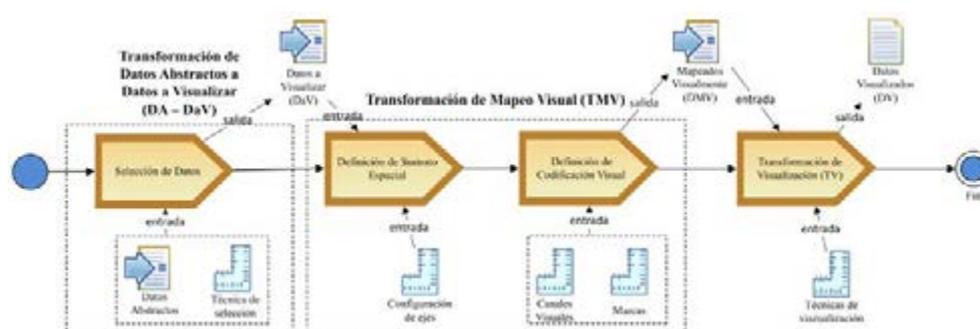
Tabla 3
Categorías por índice de masa corporal

IMC	Peso
Inferior a 18.5	Bajo peso
18.5 - 24.9	Peso saludable
25.0 - 29.9	Sobrepeso
30.0 o más	Obesidad

Modelado de Visualización

El modelado de visualización propuesto constituye un extracto del MUV, el cual consiste en un proceso interactivo entre las siguientes transformaciones: Datos Abstractos - Datos a Visualizar (DA - DaV), Mapeo Visual (TMV) y Visualización (TV), como se puede observar en la Figura 4.

Figura 4
Modelado para la visualización



A continuación, se presenta una breve descripción de las transformaciones que involucró la fase del modelado de visualización. Se tomaron como guía de referencia las investigaciones realizadas por Luján (2018) y Martig et al. (2003) en el MUV:

DA – DaV

El objetivo de esta transformación es definir el conjunto de datos que va a estar presente en la visualización y permitir la exploración de distintas regiones del espacio de información y su comparación. Para ello, se seleccionaron los datos a visualizar, y para un mismo conjunto de Datos Abstractos (DA) se generaron varios conjuntos de Datos a Visualizar (DaV), cada uno corresponde a una vista en particular.

TMV

El principal objetivo de esta transformación es conseguir un mapeo expresivo y efectivo. El mapeo es expresivo cuando se consigue representar todos los datos del conjunto de DaV. Por otra parte, la efectividad del mapeo estará definida por la manera en que la representación visual sea percibida por el usuario. En esta transformación se especificó cómo visualizar los datos presentes en el conjunto de DaV. Se generaron las estructuras necesarias con la finalidad de soportar el sustrato espacial, los elementos visuales (marcas y canales visuales) y los atributos gráficos de los elementos visuales.

El sustrato espacial está formado por la configuración entre la cantidad de ejes involucrados en la visualización, su orientación (radial, rectilínea, libre o paralela), la colección de ejes asociados y su organización en pantalla (por composición, alineación, sobrecarga o recursión). Por otro lado, el sustrato gráfico (elementos visuales y atributos gráficos) se compone de todos los elementos utilizados para representar la vista. La información presente en el sustrato gráfico se descompone en marcas (elementos visuales) y en canales visuales (controlan la apariencia de las marcas) (Luján, 2018). Según Luján (2018), para conseguir un mapeo expresivo y efectivo, es necesario establecer:

- Cuáles son las estructuras visuales adecuadas.
- Qué atributos se mapearán espacialmente y cómo.
- Qué marcas se utilizarán y mediante qué canales.

TV

Esta transformación permitió presentar en pantalla todos los datos pertenecientes al conjunto de DMV, y generar las vistas a través de la aplicación de técnicas de visualización que soporten las restricciones establecidas en el estado anterior.

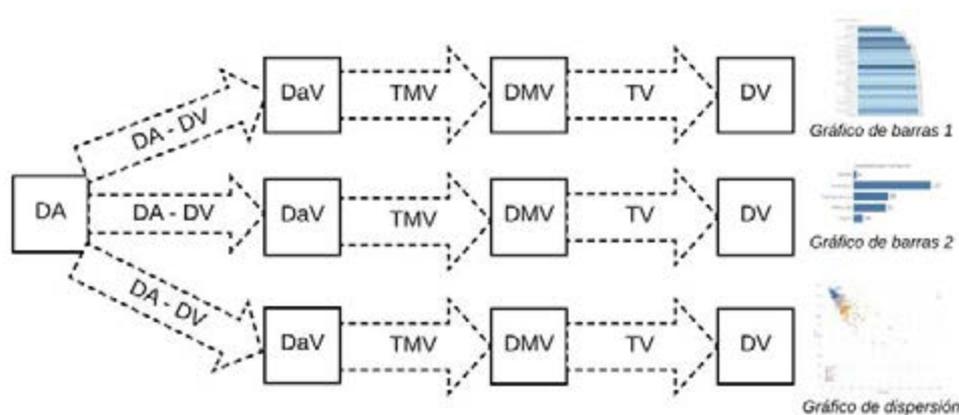
Como parte del proceso de modelado de visualización se seleccionó un conjunto de técnicas adecuadas que permitan al especialista analizar y explorar las características propias de cada jugador (demográficas, socioeconómicas y conductuales) contra aquellas variables que denotan el rendimiento cognitivo (puntaje y tiempo) en el juego serio de pares. Entre las técnicas seleccionadas se encuentran gráficos de barras, gráficos de dispersión, gráficos de líneas, gráficos de violín, diagramas de árbol, mapas de calor, etc. Para el modelado de visualización se utilizó Tableau, una plataforma de análisis integral fundada en 2013 como resultado de un proyecto de la Universidad de Stanford. Como resultado del proceso se generó una herramienta de visualización compuesta por tres tableros interactivos, *JSViz_1*, *JSViz_2* y *JSViz_3* que conforman un entorno G unificado para el análisis visual. En los siguientes apartados se detalla el proceso asociado al desarrollo de cada entorno desde la perspectiva del MUV.

JSViz_1

Es un entorno unificado de análisis visual para datos provenientes del juego serio de pares. El sistema fue diseñado para el análisis y exploración de variables demográficas y aquellas propias del juego que denotan el rendimiento de los jugadores durante cada sesión de entrenamiento. La herramienta incorpora técnicas de visualización como gráficos de barras y un diagrama de dispersión, y además provee un sistema de categorización interactivo de acuerdo con el grupo etario, así como también la clasificación de intentos según el rendimiento cognitivo. En la Figura 6 se pueden observar las transformaciones por las cuales atraviesan los datos, debido a que las tres vistas exploran el mismo conjunto de Datos a Visualizar, las representaciones correspondientes y los procesos que atraviesan en el modelo son similares hasta la transformación DA – DV. Mientras que la transformación de mapeo visual y la transformación de visualización son únicas para cada vista.

Figura 6

JSViz_1 en el MUV. Ramificaciones del modelo propuesto generando tres vistas correlacionadas



Transformación de Mapeo Visual

Se realizó la TMV para cada una de las vistas de forma independiente y como resultado del proceso se obtuvo un conjunto de Datos Mapeados Visualmente, que corresponden al estado previo a la visualización final.

- Transformación de Mapeo Visual para la vista. *Gráfico de barras 1*
 - Sustrato espacial: se definieron 2 ejes, organizados por composición, con orientación rectilínea. El eje vertical es de tipo categórico y el horizontal cuantitativo.
 - Marcas: Todos los ítems de dato se representan a través de la misma marca, un área (2D).
 - Canales: El canal visual utilizado fue el tamaño.
- Transformación de Mapeo Visual para la vista. *Gráfico de barras 2*
 - Sustrato espacial: se definieron 2 ejes, organizados por composición, con orientación rectilínea. El eje vertical es de tipo categórico y el horizontal cuantitativo.
 - Marcas: Todos los ítems de dato se representan a través de la misma marca, un área (2D).
 - Canales: Los canales visuales utilizados fueron el tamaño, que se asocia con la cantidad de puntos, y el color mediante una escala secuencial azul que representa la edad del jugador.

- Transformación de Mapeo Visual para la vista. *Gráfico de dispersión*
 - Sustrato espacial: se definieron 2 ejes, organizados por composición, con orientación rectilínea. Ambos ejes (horizontal y vertical) son de tipo cuantitativo. El rango establecido para el eje horizontal es de 0 a 600 y el vertical de 0 a 2000.
 - Marcas: Todos los ítems de dato se representan a través de la misma marca, un punto.
 - Canales: El canal visual utilizado es el color, que representa la clasificación con respecto al rendimiento al cual cada punto pertenece. El mapeo del color con el rendimiento asociado se puede observar en la Tabla 5.

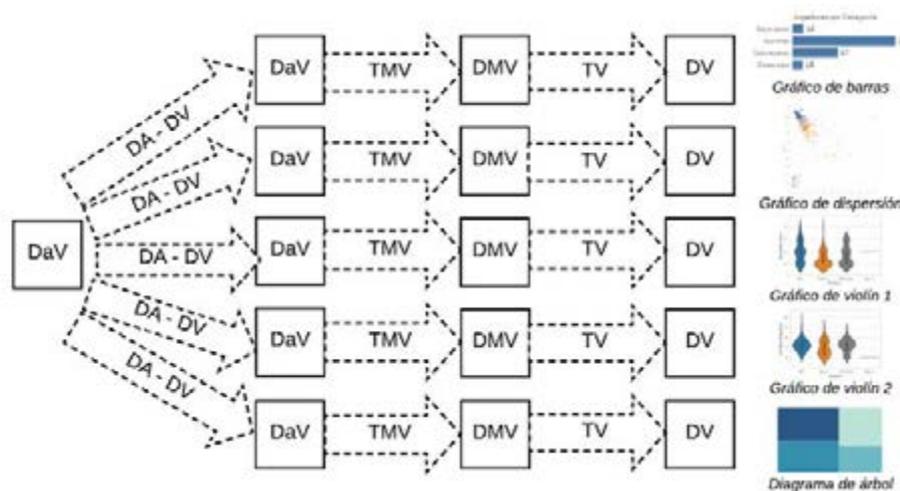
JSViz_2

Es el segundo entorno unificado de análisis visual que forma parte del modelo propuesto. A diferencia de *JSViz_1*, este entorno fue diseñado para el análisis y exploración de variables conductuales contra aquellas propias del juego que denotan el rendimiento de los jugadores durante cada sesión de entrenamiento. La herramienta incorpora técnicas de visualización como: gráficos de barras, diagrama de dispersión, diagrama de árbol y gráficos de violín, además provee un sistema de categorización interactivo de acuerdo con el Índice de Masa Corporal y la clasificación de intentos según el rendimiento cognitivo.

En la Figura 7, se pueden observar las transformaciones por las cuales atraviesan los datos y para cada vista se generó una rama. Debido a que las tres vistas exploran el mismo conjunto de Datos a Visualizar, las representaciones correspondientes y los procesos que atraviesan en el modelo son similares hasta la transformación DA – DV. Mientras que la transformación de mapeo visual y la transformación de visualización son únicas para cada vista.

Figura 7

JSViz_2 en el MUV. Configuración del proceso mediante cuatro ramificaciones para representar las vistas proporcionadas por JSViz_2



Transformación de Mapeo Visual

Se realizó la TMV para cada una de las vistas de forma independiente y como resultado del proceso se obtuvo un conjunto de Datos Mapeados Visualmente, que corresponden al estado previo a la visualización final.

- Transformación de Mapeo Visual para la vista. *Gráfico de barras*
 - Sustrato espacial: se definieron 2 ejes, organizados por composición, con

- orientación rectilínea. El eje vertical es de tipo categórico y el horizontal cuantitativo.
- Marcas: Todos los ítems de dato se representan a través de la misma marca, un área (2D).
 - Canales: El canal visual utilizado fue el tamaño.
- Transformación de Mapeo Visual para la vista. *Gráfico de dispersión*
 - Sustrato espacial: se definieron 2 ejes, organizados por composición, con orientación rectilínea. Ambos ejes (horizontal y vertical) son de tipo cuantitativo. El rango establecido para el eje horizontal es de 0 a 600 y el vertical de 0 a 2000.
 - Marcas: Todos los ítems de dato se representan a través de la misma marca, un punto.
 - Canales: El canal visual utilizado es el color, que representa la clasificación con respecto al rendimiento al cual cada punto pertenece. La codificación del color se presenta en la Tabla 5.
 - Transformación de Mapeo Visual para la vista. *Gráfico de violín 1*
 - Sustrato espacial: se definieron 2 ejes, organizados por composición, con orientación rectilínea. El eje vertical es de tipo cuantitativo y representa el tiempo de actividad física con un rango de 0 a 4 y el horizontal categórico con valores que corresponden a la clasificación por clústeres según el rendimiento: Alto, Bajo 1, Promedio y Bajo 2.
 - Marcas: Todos los ítems de dato se representan a través de la misma marca, un área (2D).
 - Canales: Los canales visuales utilizados fueron el tamaño, y el color mediante una escala categórica.
 - Transformación de Mapeo Visual para la vista. *Gráfico de violín 2*
 - Sustrato espacial: se definieron 2 ejes, organizados por composición, con orientación rectilínea. El eje vertical es de tipo cuantitativo y representa el tiempo de uso del computador con un rango de 0 a 25 y el horizontal categórico con valores que corresponden a la clasificación por clústeres según el rendimiento: Alto, Bajo 1, Promedio y Bajo 2.
 - Marcas: Todos los ítems de dato se representan a través de la misma marca, un área (2D).
 - Canales: Los canales visuales utilizados fueron el tamaño, y el color mediante una escala categórica.
 - Transformación de Mapeo Visual para la vista. *Diagrama de árbol*
 - Sustrato espacial: se definieron 2 ejes, organizados por recursión, con orientación rectilínea. El eje vertical es de tipo cuantitativo y representa el tiempo de uso del computador con un rango de 0 a 25 y el horizontal categórico con valores que corresponden a la clasificación por clústeres según el rendimiento: Alto, Bajo 1, Promedio y Bajo 2.
 - Marcas: Todos los ítems de dato se representan a través de la misma marca, un área (2D). Cada una de las marcas corresponde a la clasificación por clústeres según el rendimiento: Alto, Bajo 1, Promedio y Bajo 2.
 - Canales: Los canales visuales utilizados fueron el tamaño del rectángulo, que tiene un área proporcional a la cantidad de horas promedio de uso del computador, y el color, mediante una escala secuencial proporcional al tiempo promedio de actividad física, codificada con un matiz azul.

JSViz_3

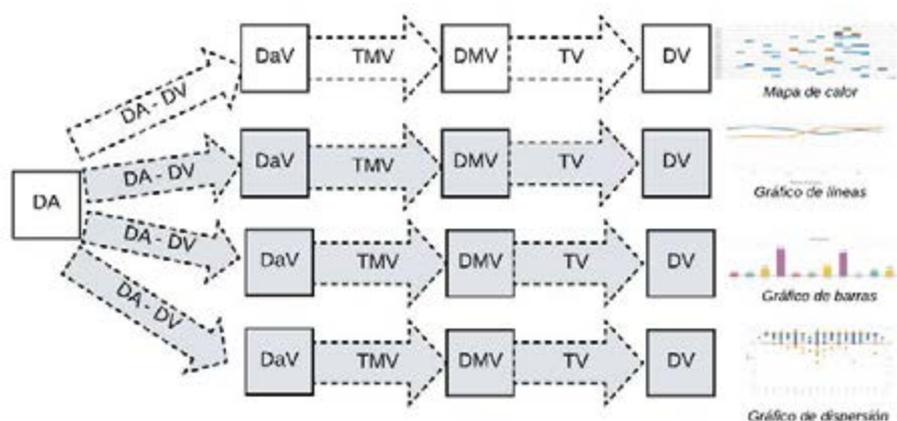
Es el tercer entorno unificado de análisis visual que forma parte del modelo propuesto.

A diferencia de los anteriores, este entorno fue diseñado para el análisis y exploración de variables temporales contra aquellas variables que denotan el rendimiento de los jugadores durante cada sesión. La herramienta incorpora técnicas de visualización como mapas de calor, gráficos de polilíneas, gráficos de barras y gráficos de dispersión, además provee un sistema de categorización interactivo de acuerdo con el número de intento, en el cual es posible comparar el rendimiento de todos jugadores e identificar valores atípicos mediante un sistema de control interactivo.

En la Figura 8 se pueden observar las transformaciones por las cuales atraviesan los datos y para cada vista se generó una rama. Al igual que en los entornos anteriores las cuatro vistas que conforman *JSViz_3* exploran el mismo conjunto de Datos a Visualizar y las representaciones correspondientes y los procesos que atraviesan en el modelo son similares hasta la transformación DA – DV. Mientras que la transformación de mapeo visual y la transformación de visualización son únicas para cada vista. Para conseguir esta visualización, se seleccionó de los DA la fecha de juego y las variables propias del juego que denotan el rendimiento de los jugadores. Los atributos seleccionados pasarán a formar parte de los DaV, descartando a todos los restantes.

Figura 8

JSViz_3 en el MUV. Las ramificaciones de color se generan de forma dinámica a partir de interacciones en el sistema de visualización



Transformación de Mapeo Visual

Se realizó la TMV para cada una de las vistas de forma independiente y como resultado del proceso se obtuvo un conjunto de Datos Mapeados Visualmente, que corresponden al estado previo a la visualización final.

- Transformación de Mapeo Visual para la vista. *Mapa de Calor*
 - Sustrato espacial: se definieron 2 ejes, organizados por composición, con orientación rectilínea. El eje vertical es de tipo categórico y el horizontal de tipo fecha.
 - Marcas: Todos los ítems de dato se representan a través de la misma marca, un área (2D). Mediante una alineación en forma de matriz, donde cada celda codifica valores cuantitativos mediante el color.
 - Canales: El canal visual utilizado fue el color, mediante una escala divergente naranja-azul, con un rango de 850 a 1975 puntos, el indicador de alerta tiene un valor de 1469, que corresponde al prototipo de un jugador con rendimiento Bajo 1, de acuerdo con la clasificación resultante del proceso de “Clasificación

Según el Rendimiento”.

- Transformación de Mapeo Visual para la vista. *Gráfico de líneas*
 - Sustrato espacial: se definieron 3 ejes, organizados por composición, con orientación rectilínea. Todos los ejes son de tipo cuantitativo. El rango establecido para el eje horizontal varía dependiendo de la cantidad de intentos registrados de cada jugador, el primer eje vertical tiene un rango establecido de 0 a 2000 y el segundo eje vertical se establece de acuerdo con el tiempo total máximo entre todos los intentos de cada jugador.
 - Marcas: Todos los ítems de dato se representan a través de la misma marca, una polilínea.
 - Canales: El primer canal visual utilizado es el color, mediante una escala categórica, que inicialmente se asignará a un color por defecto. El color naranja representa el tiempo total y el color azul el puntaje. El segundo canal visual corresponde al tamaño que varía de acuerdo con el número de intentos registrados por cada jugador.
- Transformación de Mapeo Visual para la vista. *Gráfico de barras*
 - Sustrato espacial: se definieron 2 ejes, organizados por composición, con orientación rectilínea. El eje vertical es de tipo cuantitativo y el horizontal es categórico.
 - Marcas: Todos los ítems de dato se representan a través de la misma marca, un área (2D).
 - Canales: El canal visual utilizado fue el tamaño y el color de las marcas. El color se asignará mediante una escala categórica, con colores predefinidos. El color rojo corresponde a la marca que representa el Nivel 1, el color turquesa al Nivel 2, el color amarillo al Nivel 3 y el color violeta al Nivel 4.
- Transformación de Mapeo Visual para la vista. *Gráfico de dispersión*
 - Sustrato espacial: se definieron 2 ejes, organizados por composición, con orientación rectilínea. El eje horizontal es de tipo fecha y el eje vertical es de tipo cuantitativo.
 - Marcas: Para esta vista se establecen dos tipos de marca. Un punto, que representa cada uno de los ítems del dato. Y dos líneas, que representan los límites superior e inferior de acuerdo con un parámetro calculado a partir de la desviación estándar del puntaje de todos los intentos mostrados en la vista.
 - Canales: El canal visual utilizado es el color de cada elemento visual. Las marcas que corresponden a los ítems de dato pueden tener dos colores, azul para aquellas que se encuentran dentro de los límites inferior y superior, y naranja para aquellas marcas que se encuentren fuera de los límites. De esta forma se facilita la identificación de valores atípicos.

Resultados y Discusión

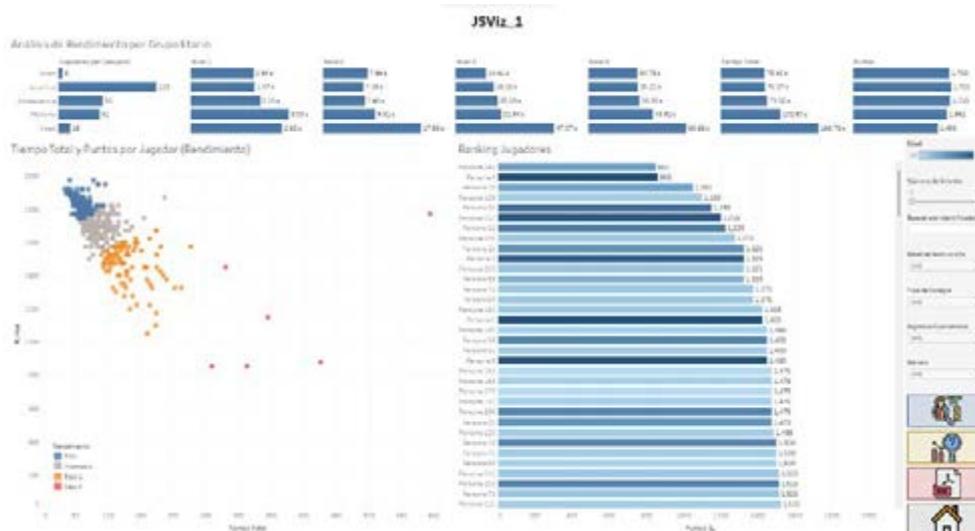
Como resultado de la metodología propuesta para el desarrollo de un modelo de visualización, se obtuvo un entorno unificado para la exploración y análisis visual de datos provenientes del juego serio de pares. El análisis a través de la visualización conduce al descubrimiento de patrones (p. ej. tendencias, brechas, valores atípicos o agrupaciones), la verificación de hipótesis y el soporte para el razonamiento y la toma de decisiones.

JSViz_1

Provee un conjunto de gráficos coordinados que permiten la exploración y análisis de las variables demográficas contra aquellas que representan el rendimiento en el juego. En la

Figura 9, se puede observar el entorno resultante.

Figura 9
Entorno de visualización unificado: JSViz_1



El entorno unificado JSViz_1 está compuesto por las siguientes vistas:

- Gráfico de barras 1: Utilizado para representar el rendimiento promedio de los jugadores según el grupo etario.
- Gráfico de dispersión: Utilizado para representar los intentos registrados según la clasificación asociada al rendimiento cognitivo.
- Gráfico de barras 2: Utilizado para representar el puntaje promedio de cada jugador.

El estado final del modelado particular de cada vista (Datos Visualizados), permitió la integración de un sistema unificado de visualización, que además de proporcionar al usuario una vista estática informativa, también permite la exploración a través herramientas (filtros, botones, etc.) que facilitan la interacción con las vistas.

Entre las interacciones que JSViz_1 provee, se cuenta con un conjunto de interacciones generales y particulares para cada vista, que permiten tareas de selección, navegación y filtrado, con el objetivo de facilitar la exploración de datos y el descubrimiento de patrones. A continuación, se describen las interacciones proporcionadas, como resultado del proceso de modelado de visualización:

- *Brushing y Linking*: JSViz_1 proporciona esta interacción que es aplicada a múltiples vistas, permitiendo seleccionar un subconjunto de datos en una vista y resaltarlos en todas las demás vistas coordinadas. Los cambios que se generen mediante una interacción a una vista se reflejan de forma automática en las vistas restantes. En particular, esta interacción se puede realizar al seleccionar uno de los intentos en el gráfico de dispersión o una de las barras del gráfico de barras 2.
- Filtro interactivo: Mediante esta interacción es posible partir los datos en segmentos, con la finalidad de enfocarse en subconjuntos de interés. En JSViz_1 esta interacción es viable mediante la selección de los siguientes elementos: 1) Categorías correspondientes a grupos etarios en el gráfico de barras 1. 2) Objetos del panel que incluyen filtros para el número de intentos, el nivel de instrucción, el tipo de

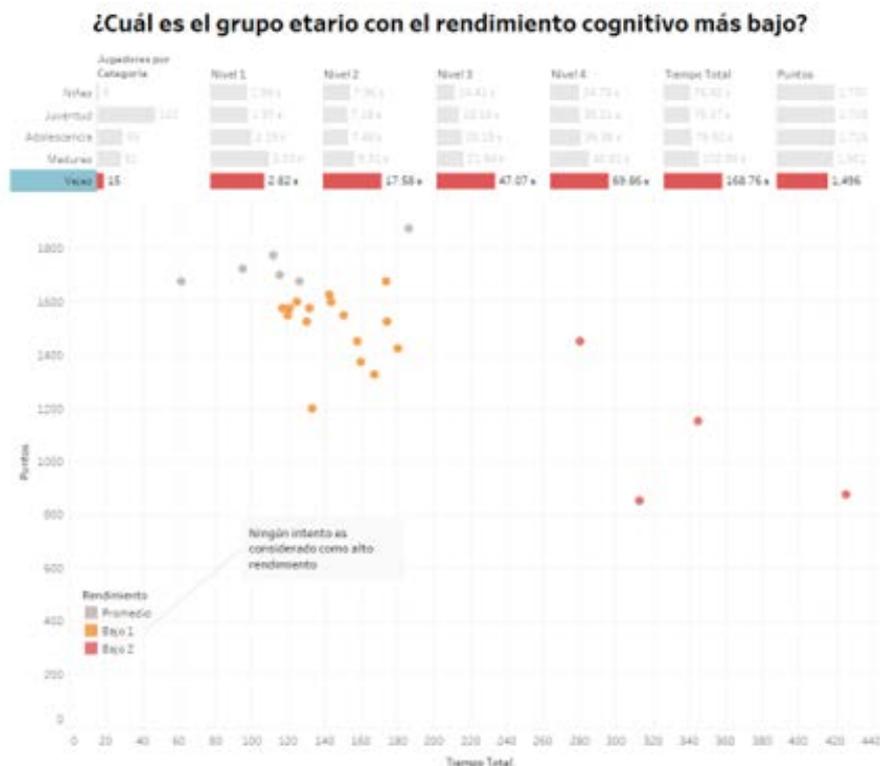
colegio, los ingresos económicos, el género y la consulta por campo de texto según el identificador del jugador.

- **Zoom Semántico:** Esta interacción permite seleccionar un elemento de la vista con la finalidad de obtener más información sobre la misma. Estos valores se muestran mediante un área de texto sobre la vista del elemento seleccionado.
- **Navegación:** Esta interacción permite navegar entre tableros. En *JSViz_1* esta interacción es posible mediante la selección de botones que se encuentran disponibles en el panel.

Las interacciones otorgan al especialista una amplia gama de combinaciones para la exploración de datos, lo que conlleva la posibilidad de responder a preguntas claves en función del campo de aplicación.

Figura 10

Ejemplo de exploración y análisis de datos con JSViz_1

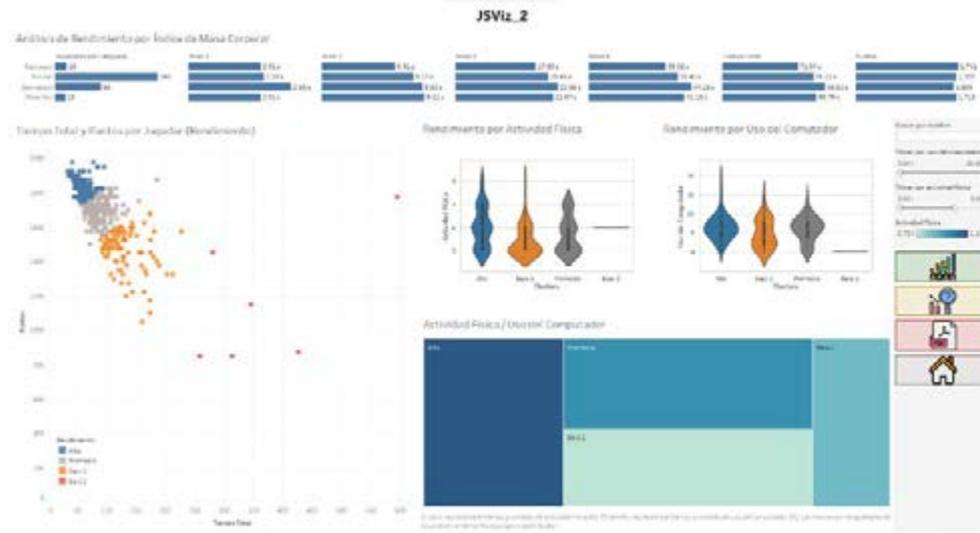


En la Figura 10 se puede observar un ejemplo de la exploración de datos utilizando *JSViz_1*, con la finalidad de responder la pregunta: *¿Cuál es el grupo etario con el rendimiento cognitivo más bajo?* A simple vista se puede deducir que el grupo etario del adulto mayor posee el rendimiento promedio más bajo del juego, en comparación a los demás grupos. También es posible determinar que ningún adulto mayor posee un desempeño asociado al grupo de alto rendimiento (marcas de color azul).

JSViz_2

Para el análisis de las variables que se asocian a la conducta o hábitos de los jugadores, se desarrolló *JSViz_2*. En la Figura 11 se puede observar el entorno resultante.

Figura 11
Entorno de visualización unificado: *JSViz_2*



El entorno unificado *JSViz_2* está compuesto por las siguientes vistas:

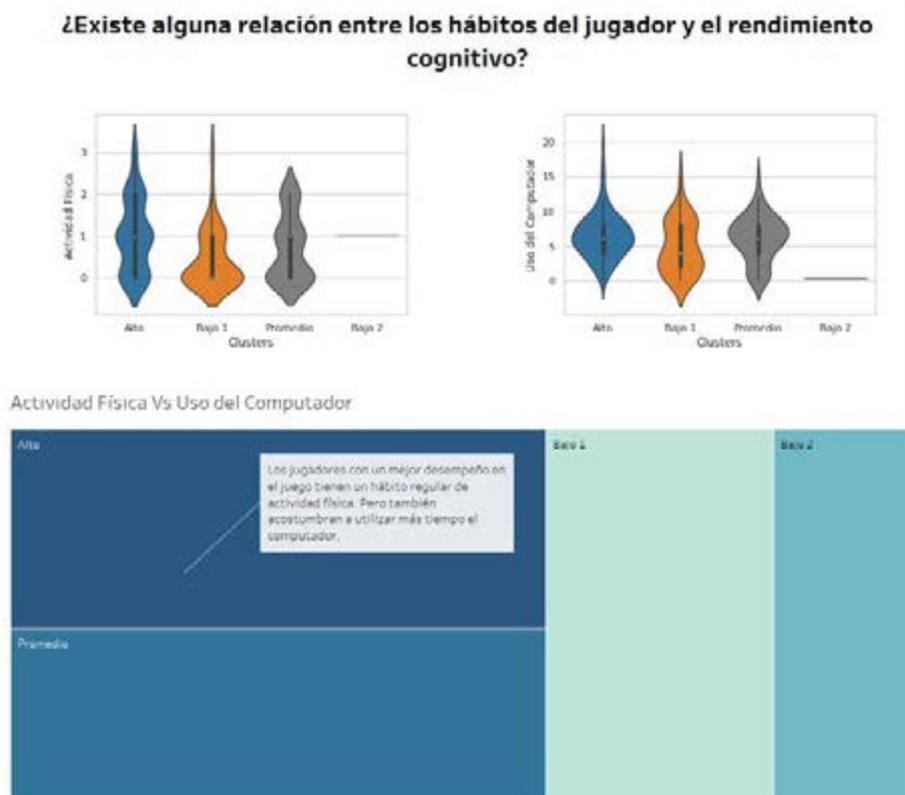
- Gráfico de barras: Utilizado para representar el rendimiento promedio de los jugadores según el Índice de Masa Corporal.
- Gráfico de violín 1: Utilizado para representar la distribución de los datos y su densidad de probabilidad asociada al tiempo de actividad física de acuerdo con la clasificación del rendimiento cognitivo y además, similar a un diagrama de cajas y bigotes, el gráfico de violín permite visualizar estos grupos de datos numéricos a través de sus cuartiles.
- Gráfico de violín 2: Utilizado para representar la distribución de los datos y su densidad de probabilidad asociada al tiempo de uso del computador de acuerdo con la clasificación del rendimiento cognitivo. Permite al usuario detectar patrones relacionados a la conducta de los jugadores según su desempeño.
- Gráfico de dispersión: Utilizado para representar los intentos registrados según la clasificación asociada al rendimiento cognitivo.
- Diagrama del árbol: Utilizado para representar la relación entre el tiempo promedio de actividad física y el tiempo promedio de uso del computador de cada categoría según el rendimiento cognitivo.

Entre las distintas interacciones que permite *JSViz_2* tenemos:

- *Brushing y Linking*: Esta interacción se puede realizar al seleccionar una de las marcas (intentos) en el *gráfico de dispersión* y como consecuencia se resalta todos los intentos relacionados al jugador en cuestión.
- Filtro interactivo: En *JSViz_2* esta interacción es posible mediante la selección de los siguientes elementos: 1) categorías correspondientes al IMC en el *gráfico de barras* 1, 2) objetos del panel que incluyen filtros para el tiempo de uso del computador y el tiempo de actividad física mediante un deslizador de rangos y la consulta por campo de texto según el identificador del jugador.
- Zoom Semántico: Esta interacción permite seleccionar un elemento de la vista con la finalidad de obtener más información sobre la misma. Estos valores se muestran mediante un área de texto que resalta sobre la vista del elemento seleccionado.
- Navegación: Esta interacción permite navegar entre tableros. En *JSViz_2* esta interacción es posible mediante la selección de botones que se encuentran disponibles en el panel.

Figura 12

Ejemplo de exploración y análisis de datos con JSViz_2



Al igual que en *JSViz_1*, es posible utilizar las interacciones disponibles con la finalidad de potenciar las capacidades cognitivas humanas y permitir el análisis de grandes volúmenes de datos a través de la visualización. En la Figura 12, se puede observar que el grupo de alto rendimiento está asociado a un hábito regular de actividad física. Mientras que los jugadores con un rendimiento inferior, en su gran mayoría no realizan actividad física.

JSViz_3

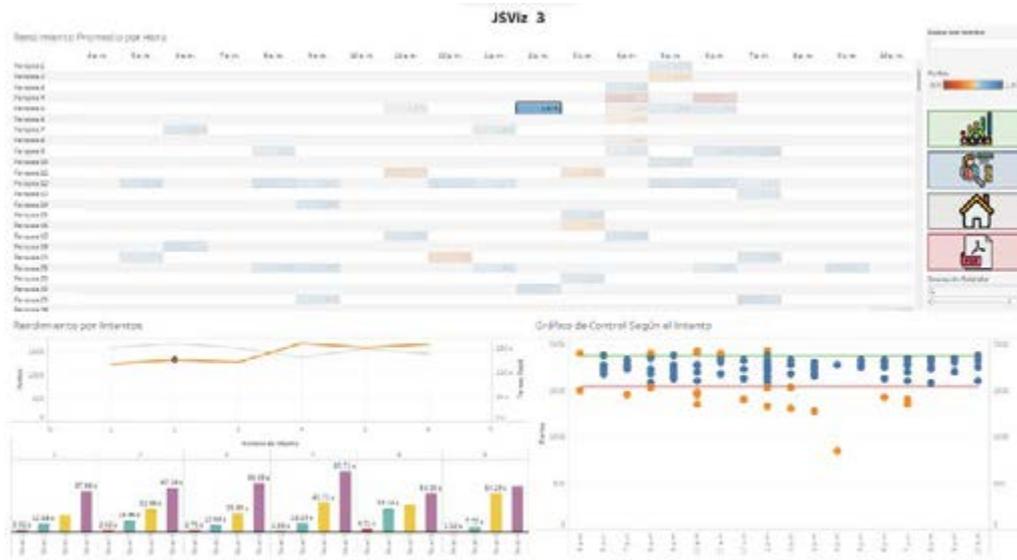
Provee un conjunto de gráficos coordinados que permiten la exploración y análisis de las variables temporales contra aquellas que representan el rendimiento en el juego. El principal objetivo de este entorno unificado es proporcionar al especialista una herramienta interactiva que facilite la evaluación de resultados con respecto a la evolución del rendimiento, a medida que se realizan las sesiones de entrenamiento. Este entorno unificado también proporciona visualizaciones diseñadas para la exploración con respecto a la hora en la que se registran sesiones y el contraste contra otros jugadores en un intento determinado, lo cual permite resaltar la presencia de valores atípicos (jugadores que denotan un desempeño estadísticamente fuera de lo común). En la Figura 13 se presenta la integración de vistas que conforman *JSViz_3*.

El entorno unificado *JSViz_3* está compuesto por las siguientes vistas:

- Mapa de calor: Utilizado para representar el rendimiento promedio de cada jugador, según la hora del juego.
- Gráfico de líneas: Utilizado para representar el rendimiento de un jugador en particular seleccionado en el mapa de calor. Los ejes verticales representan el desempeño del jugador, mientras que el eje vertical, el número de intentos.
- Gráfico de barras: Utilizado para representar el tiempo de cada nivel, agrupado por números de intentos.

- Gráfico de dispersión: Utilizado para comparar el rendimiento de los jugadores de acuerdo con un intento específico, seleccionado en el gráfico de líneas. Permite identificar valores atípicos, mediante un sistema estadístico filtrado por desviaciones estándar.

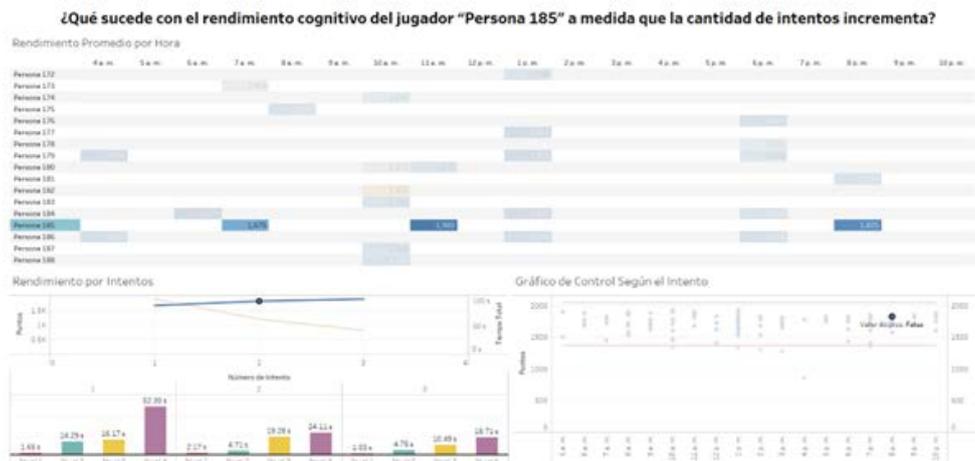
Figura 13
Entorno de visualización unificado: *JSViz_3*



Entre las distintas interacciones proporcionadas por *JSViz_3* tenemos:

- **Brushing y Linking:** Esta interacción se puede realizar al seleccionar una de las marcas en el *mapa de calor* y como consecuencia se resaltan todas las celdas relacionadas al jugador en cuestión.
- **Zoom Semántico:** Esta interacción permite seleccionar un elemento de la vista con la finalidad de obtener más información sobre la misma. Estos valores se muestran mediante un área de texto que resalta sobre la vista del elemento seleccionado.
- **Crear nueva vista:** Esta interacción es exclusiva para *JSViz_3*, permite incorporar una nueva vista al sistema de visualización y como consecuencia de la interacción se generará una nueva ramificación a partir del estado DA. En la Figura 13 se pueden observar las ramas involucradas.
- **Cerrar vista:** Con esta interacción es posible cerrar vistas abiertas en el sistema de visualización y como consecuencia de la interacción se eliminará las ramificaciones generadas a partir de la interacción “Crear nueva vista”. En la Figura 13 se pueden observar las ramas involucradas.
- **Filtro interactivo:** En *JSViz_3* esta interacción permite la reducción de ítems y es posible mediante la selección de los siguientes elementos. 1) Marcas propias de cada vista: celdas del *mapa de calor* y líneas del gráfico de líneas. 2) Objetos del panel que incluyen un filtro para el número de desviaciones estándar mediante un deslizador de rangos y una consulta por campo de texto según el identificador del jugador.
- **Navegación:** Esta interacción permite navegar entre tableros. En *JSViz_3* esta interacción es posible mediante la selección de botones que se encuentran disponibles en el panel.

Figura 14
Ejemplo de exploración y análisis de datos con JSViz_3



Además de la exploración en términos de población del conjunto de datos, también es posible analizar el rendimiento cognitivo de un jugador en particular. En la Figura 14, se puede observar un ejemplo de la exploración de datos utilizando *JSViz_3*, con la finalidad de responder la pregunta: *¿Qué sucede con el rendimiento cognitivo del jugador "Persona 185" a medida que la cantidad de intentos incrementa?* En el gráfico de líneas, se puede apreciar que el jugador demuestra un incremento en cuanto al rendimiento en términos de puntaje (línea azul) y tiempo total (línea naranja) a medida que adquiere experiencia en el juego. Así mismo, se puede observar en el *gráfico de barras* que el tiempo asociado a los niveles 3 y 4 (color amarillo y violeta respectivamente), presenta un decremento a medida que las sesiones de entrenamiento se realizan.

Conclusiones

El presente trabajo se centró en el diseño y validación de un modelo de visualización, que permite a los especialistas del área de la salud mental explorar y analizar los datos recolectados de un juego serio orientado al entrenamiento de atención y memoria, con el objetivo principal de identificar patrones asociados al deterioro cognitivo.

Para cumplir con los objetivos de este estudio, se propuso un modelo de minería de datos cuyo principio teórico se fundamenta en la metodología CRISP-DM y el Modelo Unificado de Visualización. Las etapas que forman parte del modelo consisten en tareas genéricas útiles para cualquier proyecto de minería de datos, como la comprensión del problema y la comprensión y preparación de los datos. Así como también, tareas específicas para el desarrollo del modelo de visualización, donde se realizan transformaciones que parten del tratamiento de datos abstractos hasta conseguir un conjunto de datos visualizados. Técnicamente, el proceso de modelado de visualización consistió en la selección de subconjuntos de datos a visualizar, la definición del sustrato espacial y la codificación visual. Posteriormente, con la transformación de visualización se obtuvo el resultado esperado, un conjunto de gráficos apropiados para la visualización de datos provenientes del juego de pases.

La consolidación de los datos visualizados resultó en un entorno unificado de análisis visual compuesto por tres tableros interactivos: *JSViz_1*, *JSViz_2* y *JSViz_3*. Como parte de la validación del modelo, se realizó un análisis explicativo de los datos recogidos del juego, donde participaron 248 personas. Se pudo determinar que las características de mayor influencia con

respecto al rendimiento de los jugadores son la edad, el tiempo promedio de actividad física y el tiempo de uso del computador. Específicamente, los jugadores con una costumbre habitual de actividad física demuestran mayor rendimiento con respecto a los demás jugadores; por el contrario, los jugadores con un menor desempeño cognitivo se caracterizan por pertenecer al grupo etario mayor y/o un escaso hábito de actividad física. El proceso de agrupamiento reveló la presencia de cuatro aglomerados que se categorizaron de acuerdo con el grado de rendimiento asociado a la habilidad cognitiva.

Finalmente, se llevó a cabo una evaluación empírica del modelo de visualización con 16 estudiantes de psicología, siguiendo los lineamientos de un cuasi-experimento aplicado con un modelo de transferencia tecnológica. Los resultados revelaron una recepción positiva hacia la adopción de esta tecnología. En el contexto de la salud mental, este modelo ofrece implicaciones prácticas significativas, permitiendo a los especialistas identificar patrones de deterioro cognitivo y adaptar intervenciones de acuerdo con el perfil de cada paciente. Sin embargo, es prudente considerar las limitaciones del estudio, particularmente su enfoque en un único juego y la representatividad de la muestra seleccionada. Investigaciones futuras podrían beneficiarse al ampliar la diversidad de la muestra y explorar la aplicabilidad del modelo en otros juegos serios relacionados con la salud mental.

Reconocimientos

Este trabajo ha sido financiado parcialmente por el Proyecto de Investigación Ciencia de los datos en juegos serios orientados a la atención y memoria Fase III, y forma parte de la Tesis titulada “Modelo de visualización de datos en el contexto de juegos serios orientados al entrenamiento cognitivo de atención y memoria”. Los autores desean expresar su agradecimiento al Vicerrectorado de Investigaciones de la Universidad del Azuay por el respaldo en la ejecución de proyectos de investigación.

Referencias

- Abd-Alrazaq, A., Alajlani, M., Alhuwail, D., Toro, C. T., Giannicchi, A., Ahmed, A., Makhoulf, A., & Househ, M. (2022). The Effectiveness and Safety of Serious Games for Improving Cognitive Abilities Among Elderly People With Cognitive Impairment: Systematic Review and Meta-Analysis. In *JMIR Serious Games* (Vol. 10, Issue 1). <https://doi.org/10.2196/34592>
- Alonso-Fernández, C., Calvo-Morata, A., Freire, M., Martínez-Ortiz, I., & Fernández-Manjón, B. (2019). Applications of data science to game learning analytics data: A systematic literature review. *Computers and Education*, 141(April), 103612. <https://doi.org/10.1016/j.compedu.2019.103612>
- Alonso-Fernandez, C., Calvo-Morata, A., Freire, M., Martinez-Ortiz, I., & Manjon, B. F. (2021). Data science meets standardized game learning analytics. *IEEE Global Engineering Education Conference, EDUCON*, 2021- April (February), 1546–1552. <https://doi.org/10.1109/EDUCON46332.2021.9454134>
- De Troyer, O., Helalouch, A., & Debruyne, C. (2016). Towards computer-supported self-debriefing of a serious game against cyber bullying. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 10056 LNCS(November 2020), 374–384. https://doi.org/10.1007/978-3-319-50182-6_34
- Division of Nutrition, Physical Activity, and Obesity, N. C. for C. D. P. and H. P. (2022). *Healthy Weight, Nutrition, and Physical Activity*. https://www.cdc.gov/healthyweight/assessing/bmi/adult_bmi/index.html#InterpretedAdults
- Dörner, R., Göbel, S., & Effelsberg, W. (2016). Serious games. In *Springer* (Issue 1). <https://doi.org/10.7146/pas.v33i80.111719>

- Han, J. (2012). *Data Mining Concepts and Techniques*. Elsevier Inc.
- Keim, D. A., Mansmann, F., Schneidewind, J., Thomas, J., & Ziegler, H. (2008). *Visual Analytics: Scope and Challenges BT - Visual Data Mining: Theory, Techniques and Tools for Visual Analytics* (S. J. Simoff, M. H. Böhlen, & A. Mazeika, Eds.; pp. 76–90). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-71080-6_6
- Kotu, V., & Deshpande, B. (2019). Data Science: Concepts and Practice. In *Elsevier Inc.* (Vol. 2).
- Kumar, D., & Bhardwaj, D. (2011). *Rise of Data Mining: Current and Future Application Areas Dharminder*. 8(5), 256–260.
- Loh, C. S., Sheng, Y., & Dirk, I. (2015). Serious Games Analytics. In *Serious Games Analytics*. <https://doi.org/10.1007/978-3-319-05834-4>
- Luján, M. (2018). *Interacciones en Visualización*. 247.
- Manera, V., Ben-Sadoun, G., Aalbers, T., Agopyan, H., Askenazy, F., Benoit, M., Bensamoun, D., Bourgeois, J., Bredin, J., Bremond, F., Crispim-Junior, C., David, R., Schutter, B. De, Ettore, E., Fairchild, J., Foulon, P., Gazzaley, A., Gros, A., Hun, S., ... Robert, P. (2017). Recommendations for the use of serious games in neurodegenerative disorders: 2016 Delphi Panel. *Frontiers in Physiology*, 8(JUL), 1–10. <https://doi.org/10.3389/fpsyg.2017.01243>
- Martig, S., Castro, S., Fillottrani, P., & Estévez, E. (2003). *Un Modelo Unificado de Visualización*. 881–892.
- Mezrar, S., & Bendella, F. (2022). A Systematic Review of Serious Games Relating to Cognitive Impairment and Dementia. *Journal of Digital Information Management*, 20(1), 1. <https://doi.org/10.6025/jdim/2022/20/1/01-09>
- Minović, M., Milovanović, M., Šošević, U., & Conde González, M. Á. (2015). Visualisation of student learning model in serious games. *Computers in Human Behavior*, 47, 98–107. <https://doi.org/10.1016/j.chb.2014.09.005>
- Provost, F., & Fawcett, T. (2013). *Data Science for Business*. O'Reilly Media, Inc.
- Wallner, G., Canossa, A., & El-Nasr, M. S. (2018). Introduction to the special issue on visual game analytics. *Information Visualization*, 17(3), 181–182. <https://doi.org/10.1177/1473871617722040>
- Wallner, G., & Kriglstein, S. (2013). Visualization-based analysis of gameplay data - A review of literature. *Entertainment Computing*, 4(3), 143–155. <https://doi.org/10.1016/j.entcom.2013.02.002>
- Wertsch, J. V., & Tulviste, P. (1992). L. S. Vygotsky and contemporary developmental psychology. *Developmental Psychology*, 28(4), 548–557. <https://doi.org/10.1037/0012-1649.28.4.548>
- Wirth, R., & Hipp, J. (2000). CRISP-DM: towards a standard process model for data mining. Proceedings of the Fourth International Conference on the Practical Application of Knowledge Discovery and Data Mining, 29-39. *Proceedings of the Fourth International Conference on the Practical Application of Knowledge Discovery and Data Mining*, 24959, 29–39.

Modelo Visual del Comercio Externo en Exportaciones Ecuatorianas

Visual Model of Foreign Trade in Ecuadorian Export

Dario Sebastián Medina López¹ <https://orcid.org/0009-0005-6658-3952>, Marcos Orellana¹ <https://orcid.org/0000-0002-3671-9362>, Luis Tonon-Ordóñez¹ <https://orcid.org/0000-0003-2360-9911>, Jorge Luis Zambrano-Martinez¹ <https://orcid.org/0000-0002-5339-7860>

¹Laboratorio de Investigación y Desarrollo en Informática (LIDI)
Universidad del Azuay, Cuenca, Ecuador
smedina@es.uazuay.edu.ec, marore@uazuay.edu.ec,
ltonon@uazuay.edu.ec, jorge.zambrano@uazuay.edu.ec



Esta obra está bajo una licencia internacional
Creative Commons Atribución-NoComercial 4.0

Enviado: 2023/07/14

Aceptado: 2023/08/15

Publicado: 2023/10/15

Resumen

El comercio exterior desempeña un papel fundamental en la economía de un país, involucrando una gran cantidad de transacciones y datos. Este estudio se enfoca en analizar, comprender y demostrar las exportaciones de Ecuador con el objetivo de generar conocimiento para los expertos en este campo. Para lograrlo, se desarrolló un modelo de visualización de datos que abarca el período comprendido entre 2008 y 2018. El estudio fue basado en la metodología CRISP-DM, ampliamente utilizada en proyectos de minería de datos. Esta metodología se adapta a través de cinco fases iterativas para el desarrollo del modelo de minería y visualización. Mediante la implementación de esta metodología, se logró generar un modelo visual que proporciona una herramienta efectiva para manipular las principales variables relacionadas con el comercio exterior. El resultado de este estudio fue la creación de un modelo visual adaptado a una herramienta que permite explorar y analizar las exportaciones de Ecuador. Este modelo proporciona una interfaz intuitiva para manipular y examinar las variables clave, lo que puede facilitar la toma de decisiones informadas en el ámbito del comercio exterior.

Palabras clave: Clusterización, Exportaciones, Minería de datos, Modelo de gravedad, Modelo visual, Técnicas de visualización.

Sumario: Introducción, Trabajos Relacionados, Materiales y Métodos, Resultados y Discusión y Conclusiones.

Como citar: Medina, D. S., Orellana, M., Tonon-Ordóñez, L. & Zambrano-Martinez, J. L. (2023). Modelo Visual del Comercio Externo en Exportaciones Ecuatorianas. *Revista Tecnológica - Espol*, 35(2), 143-156. Recuperado a partir de <http://www.rte.espol.edu.ec/index.php/tecnologica/article/view/1051>

Abstract

Foreign trade is fundamental to the economy as it involves many transactions and data from a country. This study analyses exports from Ecuador to generate knowledge for experts in this field. A data visualization model covering 2008 and 2018 was developed to achieve this. The study was based on the CRISP-DM methodology, widely used in data mining projects. This methodology is adapted through five iterative phases to develop the mining and visualization model. Through the implementation of this methodology, it was possible to generate a visual model that provides an effective tool to manipulate the main variables related to foreign trade. The result was the creation of a graphical model adapted into a tool to explore and analyse Ecuador's exports. This model provides an intuitive interface to manipulate and examine the main variables, which can facilitate informed decision-making in foreign trade.

Keywords: Clustering, Data mining, Exports, Gravity model, Visual model, Visualization techniques.

Introducción

A nivel mundial, con el pasar de los años, los países generan grandes cantidades de datos gracias a los avances tecnológicos, permitiendo generar datos de manera más sencilla y rápida, al igual que su almacenamiento. Esta información, se genera debido a las múltiples transacciones derivadas de todas las funciones que tienen a su cargo los estados, donde el comercio exterior representa una gran parte de estas actividades. Las exportaciones representan una de las actividades económicas más importantes de un país, la cual consiste en el envío de productos a través de fronteras. Estas transacciones de comercio exterior producen grandes cantidades de datos que son almacenados generando conocimiento informativo, no obstante, es necesario transformarlo en conocimiento productivo, que genere valor y aporte a la toma de decisiones económicas del país. El Ecuador, según el programa informático Solución Comercial Integrada Mundial (WITS), exportó 2,252 productos a 156 países en el último año (World Integrated Trade Solution, 2023). Toda la información de estas transacciones se encuentra recopilada, clasificada y almacenada en el Banco Central del Ecuador (BCE) desde el año 1990 que se encuentran disponibles para su consulta en el portal web del BCE.

Sin embargo, los mismos no disponen de medios analíticos visuales que simplifiquen y faciliten su entendimiento al momento del análisis. Esto genera que los usuarios consulten enormes archivos con tablas complejas, sin obtener un análisis visual que disminuya el trabajo de procesamiento de los datos. De igual manera, el presente estudio utiliza el modelo de gravedad que asume los flujos comerciales entre dos países al ser directamente proporcionales a ciertas variables de las economías e inversamente proporcionales a la distancia (Vásquez Bernal y Tonon Ordoñez, 2021). No obstante, es necesario analizar todos estos datos y transformarlos en conocimientos que puedan ser entendidos por los expertos como personas no inmersas en este campo de estudio, y al mismo tiempo compartirlas para generar más valor (Cairo, 2011). Por tanto, es necesario crear visualizaciones de datos que expongan y ayuden al análisis de manera clara y eficaz mediante gráficos, utilizando técnicas que combinen la estética y funcionalidad (Po et al., 2020) sobre las exportaciones ecuatorianas.

Por lo tanto, en el presente estudio se desarrolló un modelo de visualización de datos de comercio exterior que permita analizar, comprender y demostrar mediante gráficas las exportaciones del Ecuador entre los años 2010 y 2018 a través del tratamiento y procesamiento de los datos almacenados en el BCE para generar nuevo conocimiento, permitiendo que las personas puedan analizar y comprender ágilmente las exportaciones del Ecuador con sus socios

comerciales a través de esta herramienta. Asimismo, se empleó el método Cross Industry Standard Process for Data Mining (CRISP-DM) que consta de cinco fases iterativas esenciales, desde la comprensión del negocio hasta la creación del modelo visual como resultado final.

Este artículo está estructurado de la siguiente manera: La Sección 2 presenta los trabajos relacionados con métodos similares, la Sección 3 explica la metodología utilizada para realizar este estudio, la Sección 4 expone los resultados que se han obtenido al aplicar la metodología, y la Sección 5 presenta las conclusiones obtenidas, conjuntamente con los trabajos.

Trabajos Relacionados

El comercio exterior, especialmente las exportaciones, está experimentando un crecimiento cada vez mayor en Ecuador y a nivel mundial. Esto genera diariamente un gran volumen de transacciones y datos que se almacenan sin generar información para la toma de decisiones. Actualmente, los investigadores se enfrentan al desafío de representar visualmente los resultados del análisis de datos, ya que poseen amplios conocimientos en estadística y programación, pero en su mayoría carecen de conceptos adecuados en diseño y visualización (Sustaita Guerrero, 2018).

La investigación de Boar et al. (2017) realiza un análisis profundo del comercio exterior de importaciones y exportaciones utilizando datos públicos y redes complejas, llegando a identificar la relación entre las empresas distribuidoras globales y las fluctuaciones del mercado de valores. Por el contrario, Jiang et al. (2018) estudian los cambios en la dependencia del comercio exterior y la estructura comercial de la provincia de Yunnan en China mediante el análisis de los datos de comercio exterior extraídos del Banco Mundial, demostrando que el comercio exterior aumenta con el pasar del tiempo, aunque a un ritmo más lento en comparación con el crecimiento del Producto Interno Bruto (PIB) y el volumen total del comercio exterior de esta provincia.

Devyatkin et al. (2018) plantean el objetivo de revelar combinaciones de productos básicos y países asociados que experimentarían un crecimiento en el valor de las exportaciones, a través de datos abiertos del comercio internacional y la producción de las Naciones Unidas, el Fondo Monetario Internacional, bases de datos meteorológicas y modelos de regresión cuantil basados en redes neuronales.

Velichko et al. (2018) investigan sobre el comercio interno de Rusia, creando un modelo de flujo de redes múltiples utilizando la plataforma en la nube denominada aplicaciones inteligentes, control y plataforma como servicio (IACPaaS) para visualizar el análisis. Estos investigadores realizan simulaciones del tráfico comercial interregional en el Lejano Oriente de Rusia, aprovechando todos los datos disponibles de comercio y generando visualizaciones para exponer los resultados de su investigación.

Asimismo, existen estudios que vinculan el comercio exterior con técnicas de minería de datos, como el análisis realizado por Jia et al. (2019) para predecir el destino de las exportaciones de petróleo crudo de América Latina utilizando técnicas de aprendizaje automático. Estos investigadores utilizan un conjunto de datos del envío de petróleo crudo procesados con técnicas de minería de datos, como árboles de decisión, bosques aleatorios y árboles potenciados, y entrenan un modelo que permita predecir los destinos de exportación del petróleo crudo.

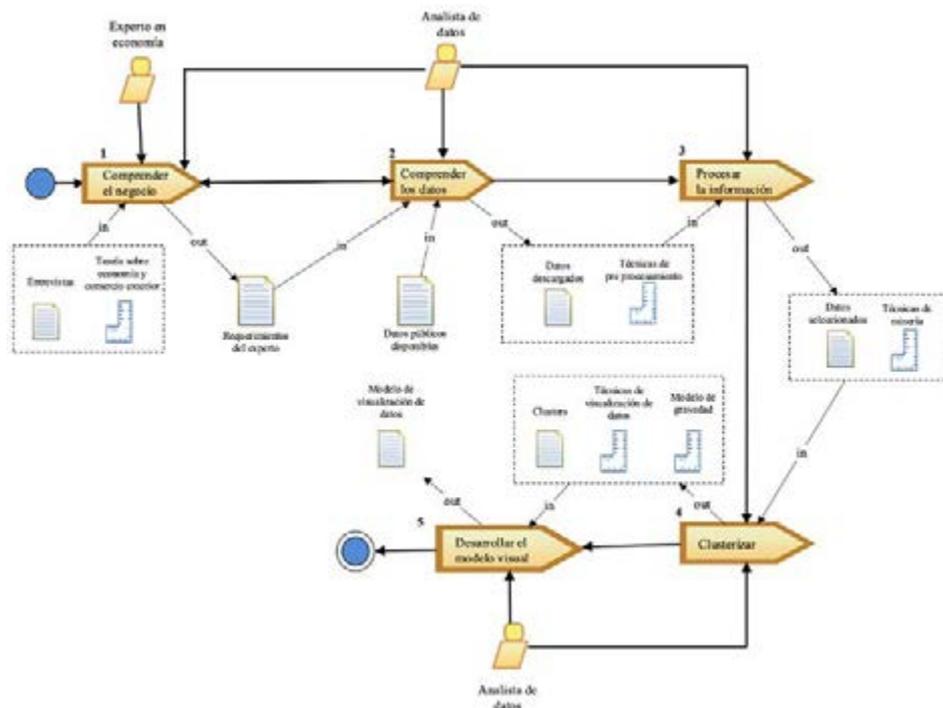
La representación visual de datos es fundamental para la toma de decisiones en el ámbito del comercio exterior, como lo evidencian Jiang et al. (2021) que usan varias gráficas y técnicas de visualización para comprender el comercio exterior a nivel mundial, identificando tendencias y comportamientos comerciales significativos.

Materiales y Métodos

El presente estudio está fundamentado en el método denominado CRISP-DM, que es ampliamente utilizado en investigaciones que involucran el tratamiento y la minería de datos. Este método se compone de cinco fases iterativas principales, que permiten un desarrollo completo del modelo de la minería de datos y la obtención de un producto final integral (IBM, 2021).

Por lo general, los investigadores que aplican este método avanzan a través de las diferentes fases del proyecto según sea necesario para mejorar la calidad de los datos y el modelo, de acuerdo con los objetivos del negocio (Saltz, 2021). CRISP-DM se basó en la metodología de descubrimiento de conocimiento en bases de datos (KDD) (Timarán Pereira et al., 2015) y se tomó como referencia para establecer la presente investigación, cuyas fases son: i) Comprender el negocio, ii) Comprender los datos, iii) Procesar la información, iv) Aplicar minería de datos (clusterizar), y v) Desarrollar el modelo visual. En la Figura 1 se presenta la metodología implementada en la investigación.

Figura 1
Metodología utilizada en la investigación



Comprender el negocio

En la primera etapa de esta investigación se realizó la comprensión del negocio, la misma que tuvo a varios involucrados como son los expertos en el área de economía, estudios internacionales y el analista de datos, así como la lectura de artículos científicos sobre el funcionamiento de exportaciones, partidas, entre otros temas fundamentales. Los expertos en economía manifestaron el apoyo que representa el poseer modelos visuales para detectar

comportamientos, cambios y tendencias dentro del comercio exterior en el Ecuador. Por lo tanto, se identificó la necesidad de desarrollar gráficos con técnicas de visualización de datos que permitan una mejor comprensión de la información del comercio exterior, y que permitan la rápida interpretación e identificación de patrones de comercio en todo el mundo. Para lograr un análisis completo fue fundamental el uso de gráficos geoespaciales y espaciotemporales a una escala mundial, caracterizando así el espacio geográfico de las exportaciones desde Ecuador hacia el mundo.

Comprender los datos

Una vez definidos los requerimientos de los expertos y la comprensión del negocio en el cual se desarrolló el proyecto para conformar el modelo visual, fue necesario comprender la información que se encuentra en las bases de datos públicas del BCE. Toda la información del BCE fue descargada desde el sitio oficial y procesada hacia archivos de formato xlsx. Estos archivos fueron clasificados por continentes e internamente cada país hace referencia a un archivo xlsx, que a su vez internamente está dividido por años. Todos los datos descargados fueron clasificados en la estructura representada en la Tabla 1, la cual hace referencia al año del registro de exportación, la cláusula de libre a bordo (FOB) en millones de dólares, la cantidad de la exportación en toneladas y el número de partida.

Tabla 1
Estructura de la información descargada desde el BCE

NOMBRE DEL CAMPO	DESCRIPCIÓN
Año	Año del registro de exportación.
FOB	Contiene sobre las partidas arancelarias.
Toneladas	Contiene las subsecciones de las partidas arancelarias.
Partida	Contiene el PIB.

Los datos recolectados fueron sobre las exportaciones de Ecuador desde 1990 hasta 2022 para su análisis. Sin embargo, debido a las circunstancias excepcionales causadas por la pandemia de COVID-19, los datos comerciales presentaban anomalías evidentes, debido a fluctuaciones de valores extremadamente anormales causadas por la pandemia. Por lo tanto, se optó por analizar los datos más recientes y libres de irregularidades causadas por situaciones excepcionales, abarcando el periodo comprendido entre 2008 y 2018.

Procesar la información

En esta fase se realizó la preparación de los datos para el modelo, garantizando su integridad y validez. Esta etapa representó entre el 50% y el 80% del esfuerzo total del proyecto. En el presente estudio, se tomaron en cuenta los archivos que contenían toda la información necesaria sobre las exportaciones de Ecuador al mundo en el periodo de 2008 a 2018. Luego, estos archivos fueron importados a una base de datos relacional, donde se organizó la información y se estructuró en tablas que permitieron realizar operaciones para depurar los datos. En la Tabla 2 se presentan los nombres y descripciones de las tablas que se han utilizado en la base de datos.

Tabla 2
Tabla de la base de datos de exportaciones.

NOMBRE DE LA TABLA	DESCRIPCIÓN
COM_CONTINENT	Contiene información de los continentes.

NOMBRE DE LA TABLA	DESCRIPCIÓN
COM_COUNTRY	Contiene información de los países.
COM_ECU_COU_DIST	Contiene la distancia de Ecuador a los países.
COMEX_EXPORT_BCE	Contiene las exportaciones de Ecuador al mundo.
COMEX_PARTIDA	Contiene sobre las partidas arancelarias.
COMEX_PARTIDA_SUB	Contiene las subsecciones de las partidas arancelarias.
COMEX_PBI	Contiene el PBI de los países.

Dado el gran volumen de información resultante y la gran cantidad de atributos en el conjunto de datos, fue necesario seleccionar aquellos atributos que brindaran más información al modelo visual y que cumplan con los objetivos del estudio.

Asimismo, se llevó a cabo un proceso de reconocimiento y limpieza de los datos mediante técnicas de preprocesamiento, con el fin de reducir el ruido y abordar posibles inconsistencias que pudieran existir en la información. Se identificaron datos nulos y aquellos que estaban fuera del rango esperado, con el objetivo de minimizar los errores que podrían surgir en la siguiente etapa. Finalmente, el conjunto de datos obtenido con todas las variables seleccionadas está representado en la Tabla 3, las cuales fueron las bases para la construcción del modelo de clustering y el modelo visual.

Tabla 3
Variables seleccionadas de exportaciones

VARIABLES	DESCRIPCIÓN
EXP_COU_ISO3	Contiene el código ISO3 del nombre del país.
EXP_YEAR	Contiene el año del registro de exportación.
PIB_VALUE	Contiene el valor del PIB del país destino
DISTFROMECU_KM	Contiene las distancias desde Ecuador.
EXP_TON	Contiene el valor de las toneladas.
EXP_DESCNAN	Contienen la descripción de la partida.
EXP_FOB	Contiene el valor de FOB.
PAR_CODIGO	Contiene el código de la partida arancelaria.
PAR_CODIGO_ECU	Contiene el código de la partida en Ecuador.
PAR_NIVEL	Contiene el nivel de la partida arancelaria.
PIB_ECU_VALUE	Contiene el valor del PIB de Ecuador.

Clusterizar

Una vez que se definió el conjunto de datos, se procedió a aplicar las técnicas de minería de datos, ya que permiten descubrir información oculta que los datos pueden contener. Para ello, es necesario conocer cuál es el flujo comercial de un país que puede ser medido por diferentes modelos, siendo el más importante el modelo de gravedad. Este modelo determina el nivel de relación comercial existente entre dos países, ya que se basa en la ley de gravitación universal de Isaac Newton que permite establecer una relación empírica entre el tamaño económico de un país y el volumen de sus importaciones y exportaciones. En la ecuación 1 se representa el modelo de gravedad de Krugman et al. (2001).

$$T_{ij} = \frac{A \times Y_i \times Y_j}{D_{ij}} \quad (1)$$

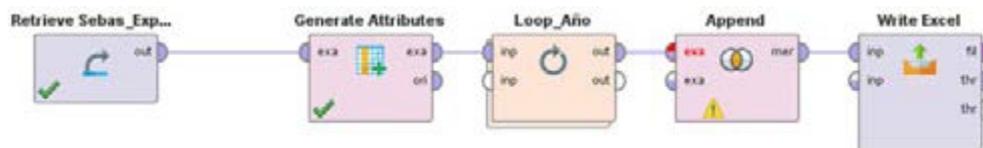
Donde A es una constante, Y_i es el PIB del país i , Y_j representa el PIB del país j y D_{ij} es la distancia entre los dos países.

Además, se utilizó la técnica de agrupamiento denominada K-means, que segmenta y agrupa la información del conjunto de datos en función de ciertos atributos similares, identificando patrones relevantes para su análisis del comercio exterior (Oyelade et al., 2019). Esta técnica fue aplicada al dataset detallado en la Tabla 3, seleccionando tres variables principales que son toneladas, FOB y PIB, y una variable más obtenida de la Ecuación 1 modelo de gravedad.

En este estudio, utilizó información correspondiente a 11 años de transacciones comerciales a diferentes niveles de partidas arancelarias, lo que dificultó la exploración de patrones. Por lo tanto, se realizó una segmentación por año y por nivel de las partidas arancelarias existentes, filtrando los datos para lograr un análisis más detallado y poder clasificarlos. Sin embargo, el conjunto de datos disponible no contiene suficientes registros para realizar el agrupamiento de las partidas de nivel uno y ocho, por lo que se omitieron en esta etapa de agrupamiento.

Figura 2

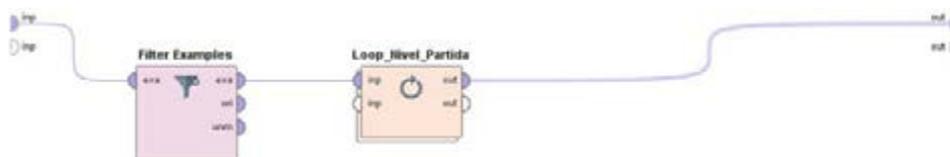
Proceso implementado en la investigación



Para aplicar la minería de datos a la información recopilada, se utilizó la herramienta RapidMiner debido a su versatilidad, eficiencia, amplia documentación y una comunidad en línea que brinda soporte. Como se observa en la Figura 2, el proceso comienza agregando el conjunto de datos resultante de la etapa anterior al software. Luego, se generó un atributo único que identifica cada registro, el cual se compone de la combinación del código del país, el código de la partida arancelaria y el año del registro, como por ejemplo "RUS20080603". A continuación, se realizaron las repeticiones del proceso descrito en los correspondientes años, para cubrir los 11 años de datos. En este proceso, se filtró por cada año de iteración y se agregó otro proceso de repetición para los niveles de las partidas arancelarias como se puede observar en la Figura 3.

Figura 3

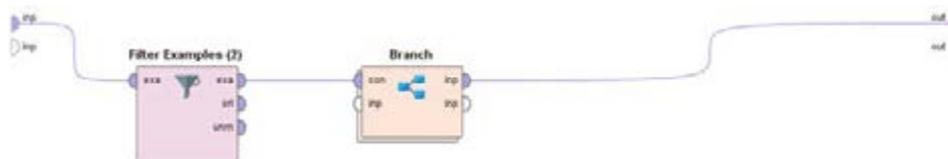
Proceso de repetición para los años



De la misma manera que pasó en la iteración por años, dentro del proceso de repetición para los niveles de las partidas arancelarias se realizó un filtrado excluyendo los niveles dos y ocho debido a la escasez de registros correspondientes a dichos niveles.

Para solucionar esta situación, se proporcionó un subproceso para bifurcaciones condicionales en función de las condiciones especificadas para las partidas arancelarias que no sean distintas de uno y ocho, como se muestra en la Figura 4.

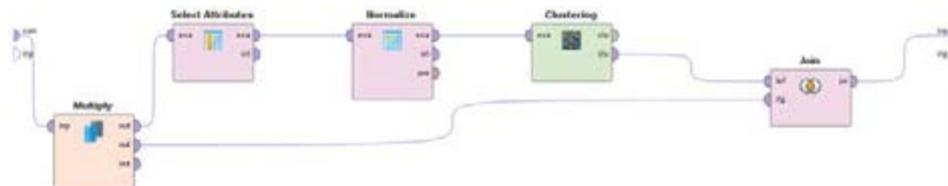
Figura 4
Proceso iterativo de los niveles de partida arancelaria



Seguidamente, como se presenta en la Figura 2, se combinaron las tablas resultantes de los modelos, generando una única tabla consolidada con la información analizada. Y, para terminar, se almacenó en un archivo el resultado del análisis que fue utilizado para la creación del modelo visual.

Para la última parte de esta fase en la metodología propuesta, se realizó el proceso de clusterización. Por lo que se realizó una duplicidad de los datos de entrada para posteriormente unirse con los resultados del clustering. Asimismo, se seleccionaron los atributos que fueron utilizados para el proceso de agrupamiento, pero fue necesario normalizar los valores de los atributos seleccionados, ajustando el rango de los valores, igualando la desviación estándar y el promedio de los atributos para realizar el análisis de agrupamiento. Y por último se realizó el agrupamiento de datos seleccionados y normalizados mediante el algoritmo k-means, como se observa en la Figura 5.

Figura 5
Proceso de clusterización de los datos



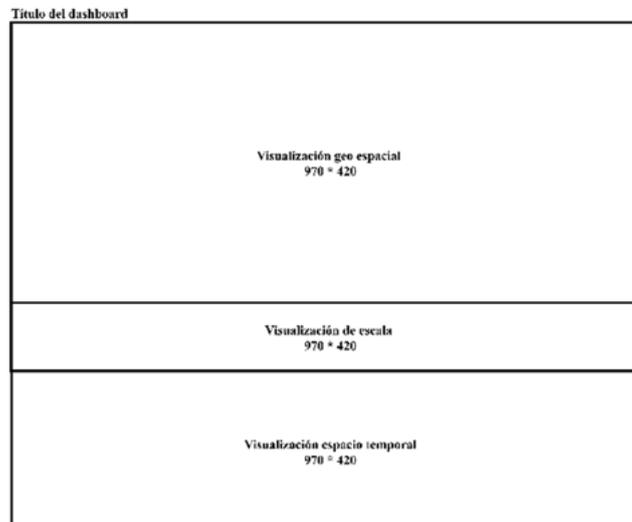
Desarrollar el modelo visual

En la quinta etapa de esta metodología, se procedió a construir el modelo visual, el cual se desarrolló teniendo en cuenta los requisitos establecidos y definidos por los expertos en la fase inicial. La necesidad de visualizar directamente los cambios en el comercio exterior resultó crucial, por lo que los gráficos principales que conformaron el modelo visual deben tener un enfoque geoespacial y espaciotemporal. Estos gráficos estaban acompañados de representaciones visuales de las variaciones, permitiendo la observación de los cambios en el comportamiento de las exportaciones a lo largo de los años. Para ello, se utilizó una paleta de colores y formas apropiadas que facilitan la detección de cambios a simple vista.

Para esta fase de la metodología se utilizó el software Tableau 2022.2 con la función de geocodificación, que genera coordenadas geográficas que pueden ser aplicadas a los mapas gratuitos de OpenStreetMap (OSM). Estas coordenadas geográficas fueron utilizadas para representar los países en el conjunto de datos después de la etapa de agrupamiento. Se empleó una paleta de colores en gradiente, del rojo al dorado, para resaltar y facilitar la comprensión de los cambios en los valores efectuados en las visualizaciones. Se utilizaron mapas coropléticos junto con la paleta de colores y las visualizaciones geoespaciales, proporcionando una representación visual detallada y en colores para el receptor.

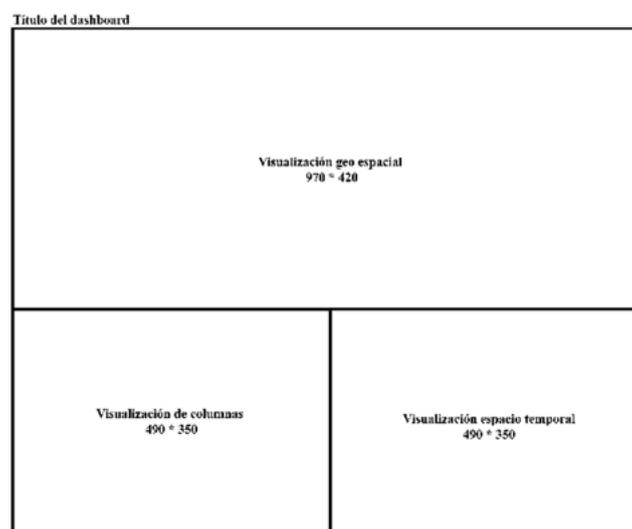
En cuanto a la disposición de las visualizaciones en el modelo visual, se creó un prototipo de interfaz de usuario en forma de maqueta, que incluye las dimensiones y la distribución precisa de las visualizaciones que estaban presentes en los paneles de resumen (dashboards). Estos paneles presentan el mensaje a transmitir a través de las visualizaciones expuestas.

Figura 6
Tablero de maquetas I



En este contexto, tal como se muestra en la Figura 6, de acuerdo con los requerimientos de los expertos, se incorporó un gráfico geoespacial como elemento central del modelo visual para facilitar la comprensión de los datos de manera efectiva. En la parte inferior, se encuentra un gráfico de barras o columnas que permite a los expertos visualizar las partidas arancelarias más frecuentes en el comercio exterior de Ecuador. Además, se incluyó un gráfico espaciotemporal que facilita el análisis de las partidas a lo largo del tiempo, proporcionando una visión clara de las variaciones entre las toneladas y el año de exportación. Es importante destacar que, en este caso, las visualizaciones del mapa y las barras permiten filtrar la información dentro de la misma visualización.

Figura 7
Tablero de maquetas II



Como se observa en la Figura 7, posee un gráfico geoespacial que representa gran parte de la visualización debido al mensaje que se desea comunicar al receptor. Luego, se ubicó una visualización tipo escala para la representación de los cambios de valores de un determinado atributo en contraste con la visualización geoespacial. Finalmente, se visualizó un gráfico de espacio temporal que representó las variaciones existentes del atributo seleccionado. De este modo, el único filtro en este tablero fue la visualización geoespacial.

Figura 8
Tablero de maquetas III



Y, por último, se incluyó una visualización tipo radar, que en conjunto con la visualización geoespacial conformó la parte principal de este tablero, como se observa en la Figura 8. En la parte inferior, existen dos filtros de selección única para la depuración de la información y una escala para intervalos de valores. Adicionalmente, existen dos visualizaciones del mismo tamaño que representan un gráfico de barras de valores promedio filtrados y otra de la paleta de colores del panel, que ayuda a la clasificación.

Resultados y Discusión

Los primeros resultados obtenidos en esta investigación se derivan de la etapa de clusterización, donde se generaron un total de 66 modelos de clustering que fueron utilizados en la creación del modelo visual. Cada uno de estos modelos cuenta con una tabla interna que resume la información más relevante de cada clúster. Así, en la Tabla 4, se muestran los cinco clústeres obtenidos para el año 2010 y el segundo nivel de partida arancelaria, junto con el número de elementos y los centros correspondientes a cada clúster.

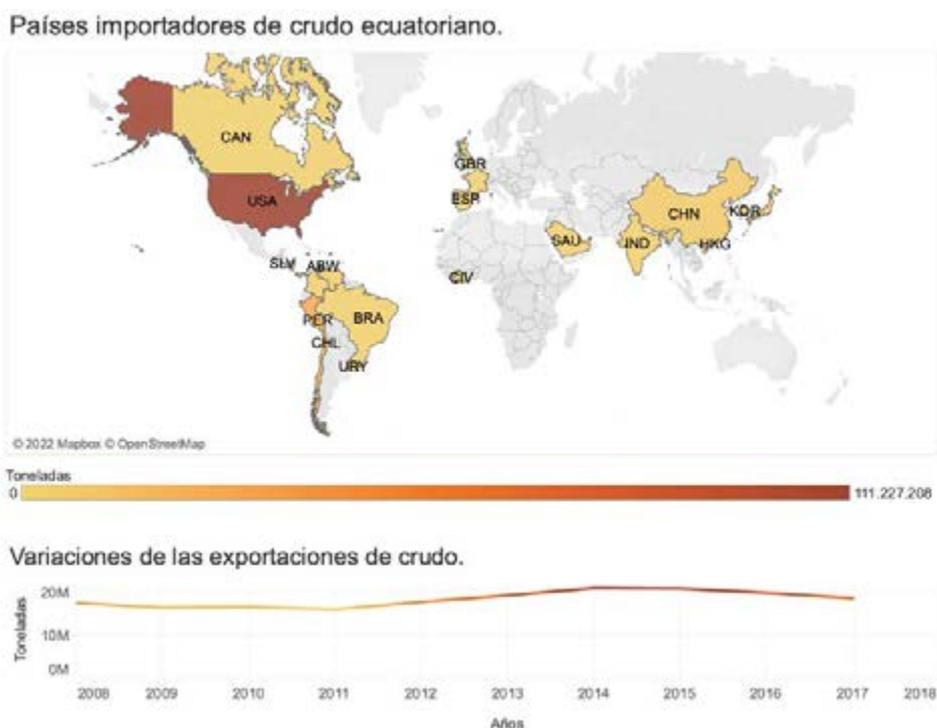
Tabla 4
Clústeres para el año 2010 con partida arancelaria nivel 2

CLÚSTERES	NÚMEROS DE ELEMENTOS	CENTRO DEL CLÚSTER
1	168	0.908
2	2	0.011
3	12	0.011
4	7	0.038
5	6	0.032

El modelo visual constituye la parte central de los resultados de esta investigación, facilitando el análisis de las exportaciones ecuatorianas. Para lo cual, en la Figura 9, se presenta el panel de control que representa las exportaciones de crudo más significativas de Ecuador hacia el resto del mundo. Este panel permite al usuario final filtrar los datos mediante un mapa mundial interactivo, que puede ser visualizado a través de coropletas. Además, con este gráfico espaciotemporal muestra las variaciones en tiempo real de las exportaciones de crudo a lo largo de los años. Cabe destacar que se utilizó una escala de colores para resaltar los países que importan una mayor cantidad de crudo, con tonalidades más intensas de rojo, mientras que los tonos más amarillos se asignan a los países con menor importancia. Esto permitió tener una visualización clara de los principales importadores y los cambios notables que se produjeron a lo largo del tiempo.

Figura 9

Modelo visual de los países importadores del petróleo



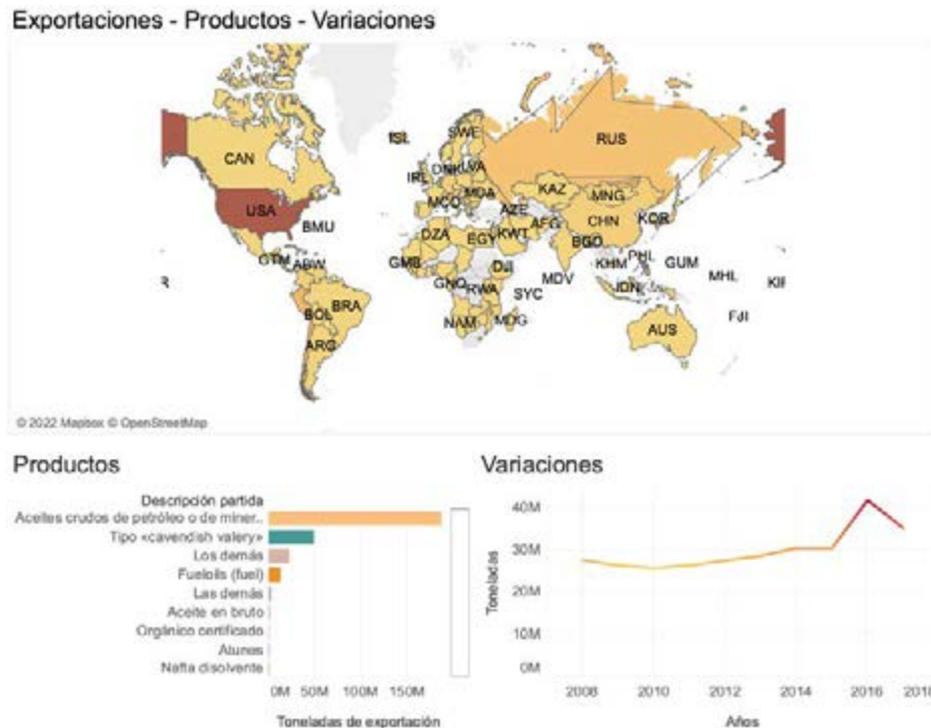
En el segundo panel de control, el usuario final puede analizar las variables más relevantes del comercio exterior de Ecuador. En este panel, se combinaron características como las exportaciones por países, productos y las variaciones existentes para lograr un análisis exhaustivo mediante el modelo visual de manera interactiva con el mapa mundial, como se presenta en la Figura 10. Por consiguiente, el modelo visual proporciona la descripción de las partidas arancelarias más utilizadas por el país seleccionado a través de un gráfico de barras ordenadas descendientemente. Por último, se puede seleccionar una partida arancelaria específica y visualizar las variaciones en las toneladas exportadas a lo largo del tiempo mediante una representación espacio-temporal.

Por último, en el tercer panel se tiene al algoritmo de minería de datos, denominado "Clustering", que a través del conjunto de visualizaciones permite realizar un análisis integral de las exportaciones. De esta manera, se examinó de forma indirecta el FOB, PIB, toneladas y la fuerza del comercio internacional por medio del modelo de gravedad de cada país. Por lo tanto, se seleccionó una partida arancelaria y un año específico para el análisis; acto seguido, el

panel presentó los clústeres existentes y los valores de cada atributo mediante la visualización de un gráfico de radar. Además, se brindó la opción de seleccionar un clúster en particular para filtrar las visualizaciones del mapa y el gráfico de barras, proporcionando información sobre los promedios de atributos correspondientes, como se muestra en la Figura 11.

Figura 10

Modelo visual de las exportaciones, productos y variaciones



La evaluación de la fuerza del comercio internacional requiere en primera instancia la aplicación del modelo de gravedad, utilizando variables adicionales como el riesgo nacional, la inseguridad y las tasas de educación, entre otras. Estos factores mejoran la precisión de la fuerza del comercio y ofrecen un enfoque adicional para realizar un análisis meticuloso del mercado de exportación. Aunque se utiliza una de las técnicas de minería de datos más relevantes, el análisis de las exportaciones puede ampliarse mediante el uso de otras técnicas bien establecidas como los árboles de decisión y la predicción, lo que permitiría obtener conclusiones diferentes según los objetivos de los expertos.

Esta investigación demostró claramente que existe un consenso en el uso de datos de comercio exterior extraídos de fuentes públicas, coincidiendo con varios autores como Devyatkin et al. (2018), Jiang et al. (2021) y Velichko et al. (2018), quienes resaltan la importancia y utilidad del análisis de los datos generados por las transacciones comerciales, aportando valor a los datos y contribuyendo a la toma de decisiones. Además, se determinó que muchos autores no utilizan principalmente técnicas de visualización en combinación con técnicas de minería de datos, lo que permite un análisis integral de la información pública disponible de cada país en relación con el comercio exterior. En este sentido, se destaca el trabajo de Hui (2018), quien realiza un análisis profundo de los datos comerciales para la aplicación de estrategias de marketing, pero no presenta los resultados a través de visualizaciones de datos, sino a través de resúmenes en forma de tablas. Varios autores, como Boar et al. (2017) y M. Jiang et al. (2018) respaldan la importancia de la visualización de datos para comunicar los resultados de manera efectiva y facilitar su comprensión y difusión. Aunque en el estudio actual se seleccionó un amplio rango

detallado del mercado de exportación. Además, se puede ampliar el análisis de las exportaciones aplicando técnicas conocidas como los árboles de decisión y la previsión con la finalidad de estimar proyecciones de resultados posibles desde diferentes puntos de vista acordes a los objetivos de los expertos.

Reconocimientos

Los autores desean agradecer al Vicerrectorado de Investigaciones de la Universidad del Azuay por el apoyo financiero y académico, así como a todo el personal de la escuela de Ingeniería de Ciencias de la Computación, y el Laboratorio de Investigación y Desarrollo en Informática (LIDI).

Referencias

- Boar, R., Iovanovici, A., y Ciocarlie, H. (2017). Complex networks analysis of international import-export trade. *2017 IEEE 14th International Scientific Conference on Informatics*, 31-34. <https://doi.org/10.1109/INFORMATICS.2017.8327217>
- Cairo, A. (2011). *El arte de lo funcional: infografía y visualización de información*. Alamut.
- Devyatkin, D., Suvorov, R., Tikhomirov, I., y Otmakhova, Y. (2018). Neural Networks for Food Export Gain Forecasting. *2018 International Conference on Intelligent Systems (IS)*, 312-317. <https://doi.org/10.1109/IS.2018.8710561>
- Hui, Z. (2018). Analysis of International Marketing Strategy Based on Intelligent Mining Algorithm for Big Data. *2018 11th International Conference on Intelligent Computation Technology and Automation (ICICTA)*, 261-264. <https://doi.org/10.1109/ICICTA.2018.00065>
- Jiang, H., Shen, J., Chou, Q., Dong, Z., y Cheng, S. (2021). Visual Analytics for the International Trade. *2021 5th International Conference on Vision, Image and Signal Processing (ICVISP)*, 296-301. <https://doi.org/10.1109/ICVISP54630.2021.00059>
- Jiang, M., Niu, L., Zhang, Y., Wang, Z., y Ren, X. (2018). A Big-Data-Analysis-Based Research Study on the Changes Both in the Dependence on Foreign Trade and in the Trade Structure (2000-2015) of the Yunnan Province of China. *2018 International Conference on Intelligent Transportation, Big Data y Smart City (ICITBS)*, 235-239. <https://doi.org/10.1109/ICITBS.2018.00068>
- Krugman, P., Obstfeld, M., y Melitz, M. J. (2001). *Economía internacional: Teoría y política*. Pearson education Madrid.
- OpenStreetMap. (s. f.). *OSM*. Recuperado 26 de junio de 2023, de <https://www.openstreetmap.org/#map=7/-1.783/-78.132>
- Oyelade, J., Isewon, I., Oladipupo, O., Emebo, O., Omogbadegun, Z., Aromolaran, O., Uwoghiren, E., Olaniyan, D., y Olawole, O. (2019). Data Clustering: Algorithms and Its Applications. *2019 19th International Conference on Computational Science and Its Applications (ICCSA)*, 71-81. <https://doi.org/10.1109/ICCSA.2019.000-1>
- Po, L., Bikakis, N., Desimoni, F., y Papastefanatos, G. (2020). Linked data visualization: techniques, tools, and big data. *Synthesis Lectures on Semantic Web: Theory and Technology*, 10(1), 1-157.
- Vásquez Bernal, J. V, y Tonon Ordóñez, L. B. (2021). Modelo de gravedad de las exportaciones de cacao en grano del Ecuador. *INNOVA Research Journal*.
- Velichko, A., Gribova, V., y Fedorishchev, L. (2018). Simulation Software for Multicommodity Flows Model of Interregional Trade. *2018 3rd Russian-Pacific Conference on Computer Technology and Applications (RPC)*, 1-5. <https://doi.org/10.1109/RPC.2018.8482140>

Análisis de Clusterización en Datos de Encuestas sobre Ciberacoso

Clustering Analysis on Cyberbullying Survey Data

César Vinueza-Álvarez¹ <https://orcid.org/0009-0001-6731-3314>, María Inés Acosta-Urigüen¹ <https://orcid.org/0000-0003-4865-2983>, Juan Fernando Lima¹ <https://orcid.org/0000-0003-3500-3968>

¹*Departamento de Posgrados, Universidad del Azuay, Cuenca, Ecuador*
cvinuezamsn@es.uazuay.edu.ec

²*Laboratorio de Investigación y Desarrollo en Informática (LIDI), Universidad del Azuay, Cuenca, Ecuador*
macosta@uazuay.edu.ec, flima@uazuay.edu.ec



Esta obra está bajo una licencia internacional
Creative Commons Atribución-NoComercial 4.0

Enviado: 2023/07/15

Aceptado: 2023/08/22

Publicado: 2023/10/15

Resumen

Hoy en día, el ciberacoso se ha incrementado debido al crecimiento y diversificación de tecnologías, resultando difícil una detección temprana en distintos sectores, como en la academia. Por este motivo, esta problemática requiere abordarse mediante métodos específicos de análisis de datos que permitan identificar su caracterización. El presente estudio se enfoca en la detección del ciberacoso en los estudiantes de una universidad de la ciudad de Cuenca, Ecuador, aplicando minería de datos para analizar la información obtenida de un cuestionario psicológico sobre el ciberacoso con aspectos sociodemográficos, y a través de la escala de Likert llegar a categorizar el nivel de victimización y de agresión entre adolescentes al momento de utilizar dispositivos electrónicos. Mediante el uso del algoritmo de agrupamiento se evalúa y se obtienen patrones de comportamiento de los participantes de la encuesta del ciberacoso donde se identificaron a los actores y los sectores académicos con mayor incidencia.

Palabras clave: Ciberacoso, Entorno académico, Victimización encuesta, Agrupamiento.

Sumario: Introducción, Trabajos Relacionados, Materiales y Métodos, Resultados y Conclusiones.

Como citar: Vinueza-Álvarez, C., Acosta-Urigüen, M. I. & Lima, J. F. (2023). Análisis de Clusterización en Datos de Encuestas sobre Ciberacoso. *Revista Tecnológica - Espol*, 35(2), 157-169.

Recuperado a partir de <http://www.rte.espol.edu.ec/index.php/tecnologica/article/view/1055>

Abstract

Nowadays, cyberbullying has increased due to the growth and diversification of technologies. This has made early detection difficult in different sectors, such as academia. For this reason, this issue requires addressing through specific data analysis methods that allow its characterization to be identified. This study focuses on the detection of cyberbullying in students of a university in the city of Cuenca, Ecuador, applying data mining to analyze the information obtained from a psychological questionnaire on cyberbullying with sociodemographic aspects and through the scale of Likert to categorize the level of victimization and aggression among adolescents when using electronic devices. Using the grouping algorithm, the behavior patterns of the participants in the cyberbullying survey are evaluated and obtained, and the actors and academic sectors with the highest incidence are identified.

Keywords: Cyberbullying, Academic domain, Victimization survey, Clustering.

Introducción

El acoso, también conocido como “bullying”, ha sido uno de los temas que más se presentan a nivel académico, preocupando a los educadores, psicólogos e incluso a los padres de familia, debido al incremento del acoso sin presentar una distinción de edad, nacionalidad o género. Por lo tanto, llega a afectar a la sociedad trayendo consecuencias negativas en todo ámbito donde se presente (Garmendia Larrañaga et al., 2019).

El avance constante de la tecnología en el mundo moderno, así como la difusión y uso de las redes sociales, facilita el acoso constante a las víctimas por parte de los atacantes. Siendo así el surgimiento de un nuevo término denominado ciberacoso o “cyberbullying”, que no es más que trasladar el acoso físico a los distintos medios digitales (Hosseinmardi et al., 2014). En general, el ciberacoso es considerado una amenaza psicológica para los niños y adolescentes, que hacen uso de dispositivos electrónicos, debido a que pueden afectar negativamente aspectos psicológicos, conductuales, físicos y académicos en la víctima. Asimismo, se presentan ansiedad, autoestima baja, dependencia del alcohol y drogas, reducción de autocontrol, pensamientos suicidas y pésimos resultados académicos tanto en el atacante como en la víctima (Johnson, 2022).

Por este motivo, es necesario encontrar distintas alternativas para que las autoridades competentes puedan identificar a las víctimas del ciberacoso, que son constantemente acosadas por los atacantes (Moya-Solís & Moreta-Herrera, 2022). Para lo cual, la minería de datos es una alternativa robusta para encontrar patrones de comportamiento y llegar a reducir el ciberacoso con la identificación de los actores, mediante el uso de distintas técnicas como clasificación, asociación, clasificador bayesiano ingenuo (*Naive Bayes*), árbol de decisión, bosques aleatorios (Random Forest), entre otras (Pamuji & Setiawan, 2022). Por eso, para esta investigación, se propone el análisis de los datos provenientes de un cuestionario de ciberacoso mediante el algoritmo de Partición Alrededor de Medoids (PAM) que minimiza el resultado de la suma de las diferencias entre los objetos de un grupo y el centro de ese grupo (Botyarov & Miller, 2022). Para lograr esta propuesta, es necesario analizar las respuestas del cuestionario de ciberacoso mediante el algoritmo PAM, para llegar a detectar aspectos característicos donde se presentan los ciberacosos para los estudiantes universitarios de distintas carreras.

El resto del artículo se estructura de la siguiente manera: la Sección 2 presenta los trabajos relacionados; la metodología del trabajo es presentada en la Sección 3; en la Sección 4 se muestran los resultados; y finalmente, en la Sección 5 se presentan las conclusiones y los trabajos futuros.

Trabajos Relacionados

Existen estudios donde se presentan los patrones que conducen al acoso y al ciberacoso, que conducen a la generación de modelos para solucionar esta problemática social. En consecuencia, los principales desarrollos de las investigaciones se centran en la detección, clasificación y predicción del ciberacoso en diferentes sectores sociales. A continuación, se presentan los trabajos relacionados a las técnicas utilizadas para la detección del acoso y el ciberacoso.

Los autores Sitnik-Warchulska et al. (2021) demuestran que utilizar redes bayesianas llega a modelar comportamientos de riesgo de acoso entre los escolares, así como la identificación de los actores que están involucrados. Por lo tanto, usar redes bayesianas que permitan visualizar y modelar las relaciones entre distintas hipótesis y variables, para cuantificar la incertidumbre en las inferencias basadas en el análisis de datos estadísticos, permite predecir conductas de acoso. Además, Sitnik-Warchulska et al. clasifican y predicen los estados y eventos del acoso aun cuando los datos son parciales e inciertos.

Mientras tanto, Meliana et al. (2019) identifican el ciberacoso en las conversaciones estudiadas de la red social X (Twitter) mediante los métodos de agrupación. Estos autores usan la minería de datos en esta red social para obtener los resultados y su correlación con la identificación del ciberacoso. Además, en este estudio se utilizó el método de clasificación Naive Bayes y árboles de decisión (*Decision Tree J48*), los cuales identificaron exitosamente un 92% y un 100% respectivamente. Asimismo, los autores concluyen que la red social que presenta un alto ciberacoso es Twitter, relacionado con acosos psicológicos y violentos.

Por otro lado, Farag et al. (2019), en un estudio realizado, concluyen que es necesario utilizar diferentes métodos de análisis de datos para cada una de las redes sociales, debido a cómo entregan los datos estas plataformas. Por lo tanto, proponen utilizar seis métodos para analizar los datos recolectados de las redes sociales, como aprendizaje no supervisado, codificadores automáticos, aprendizaje profundo, aprendizaje semi supervisado, modelado de series de tiempo y clusterización. Además, señalan que los algoritmos de aprendizaje no supervisados están ganando una mayor atención por parte de la comunidad científica en este tema, debido a presentar un alto grado de eficiencia al momento de analizar los datos.

Los autores Parime & Suri (2014) analizan el ciberacoso, cuáles son sus impactos en la sociedad y sus principales características, por lo que utilizan técnicas de minería de datos para entender el comportamiento que presentan los acosadores y así prevenir aquellos actos. Por lo tanto, Parime & Suri analizan los sentimientos a través del procesamiento del lenguaje natural para identificar la presencia o ausencia del ciberacoso en una red social, a través de un conjunto de datos.

Mientras que, en otro estudio, realizado por Zhang (2021), se propone un método que permite identificar la presencia de términos de ciberacoso y clasificar las actividades de acoso en las redes sociales como Flaming. Además, expone cuáles son los procesos para detectarlo, tales como el pre-procesamiento de los datos, extracción de características, algoritmos de aprendizaje y técnicas de clasificación Naive Bayes. De esta manera, es posible detectar el ciberacoso en las redes sociales con la lógica difusa que proporciona una ayuda para tomar decisiones sobre los contenidos antes que los usuarios sean víctimas.

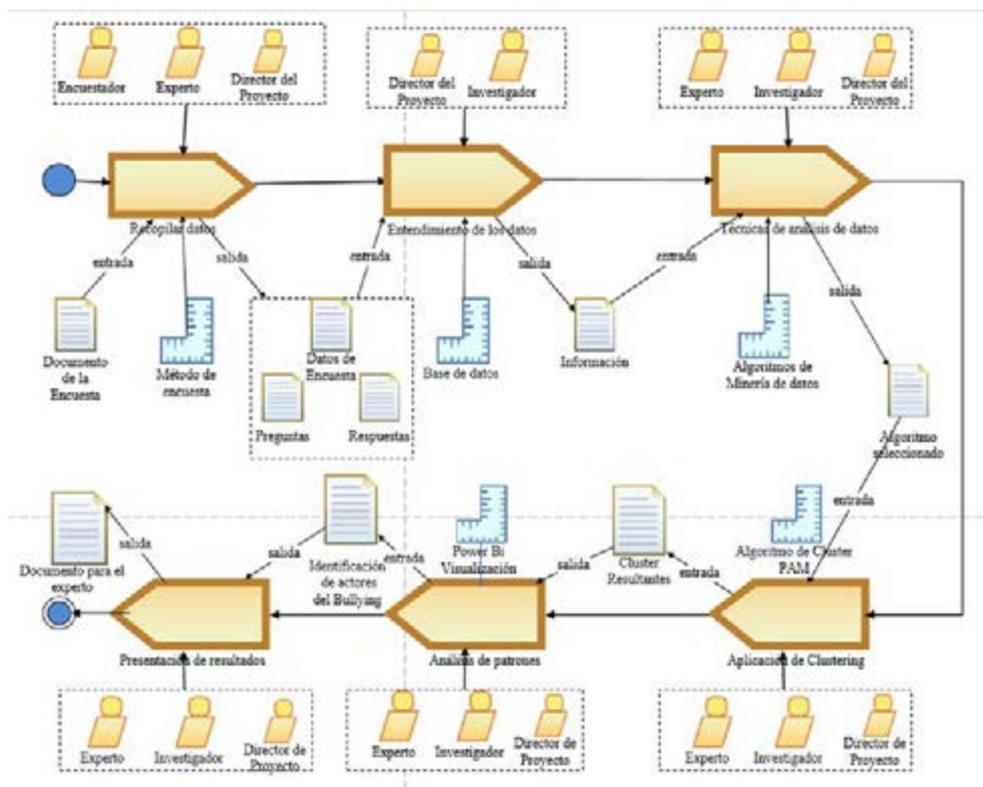
Los autores Dedic et al. (2019) analizan los datos obtenidos con tres tipos de encuestas estandarizadas (“*The Juvenile Victimization Questionnaire*”, “*The Defense Style Questionnaire*” y “*Beck Depression Inventory*”) realizadas a estudiantes del Sur de Australia, Gran Bretaña y Suecia a través de una prueba no paramétrica denominada Kolmogorov-Smirnov que determina la bondad del ajuste de dos distribuciones de probabilidad entre sí. Dedic et al. llegan a detectar los factores primordiales para cometer suicidios ocasionados en la niñez por haber sido víctimas de abusos sexuales de los compañeros.

Los estudios mencionados describen varias técnicas y modelos para la detección y predicción tanto del acoso como del ciberacoso. Presentan limitaciones al momento de detectar y predecir esta problemática ya que, si bien los estudios tienen un enfoque psicológico para el acoso y el ciberacoso, no llegan a la aplicación de técnicas de clusterización de los datos proporcionados con las distintas pruebas psicológicas.

Materiales y Métodos

La metodología empleada durante el proceso de aplicación de técnicas de clusterización de datos de una encuesta sobre el ciberacoso estuvo conformada por seis actividades detalladas en la Figura 1. Para estas actividades se utilizaron como entrada los datos provenientes de las encuestas psicológicas sobre el acoso y el ciberacoso que fueron realizadas.

Figura 1
Actividades relacionadas a la metodología utilizada en el estudio



Recopilación de los datos

Esta fase inicial está primordialmente enfocada en la comprensión del acoso y el ciberacoso como un problema social actual, donde la implementación de una encuesta psicológica permite una clara detección, a través de los datos que esta herramienta brinda. Los roles que se presentan en esta fase, acorde a la Figura 1, son:

- Encuestador: aquella persona que está encargada de aplicar los cuestionarios, escuchar, plantear las preguntas y registrar las respuestas de las personas entrevistadas.
- Experto: persona que tiene gran experiencia en el ámbito psicológico y brinda soporte a la estructura de la encuesta.
- Director del proyecto: aquella persona que dirige el estudio brinda apoyo y conocimiento de técnicas y métodos para evaluar la información y a su vez descartar la que no sea confiable.
- Investigador: es quien está encargado de la construcción y definición de cada una de las fases de la investigación que permitan una correcta gestión del proyecto.

Estudio y comprensión de los datos

En esta fase se realizó la recopilación inicial de los datos y las tareas de tratamiento de la información obtenida de las encuestas. Se llevó a cabo la identificación de las características de los datos que puedan ser de interés para el estudio, que a simple vista no se pueden observar. Por lo tanto, en la fase de entendimiento de datos se focalizó en el análisis de las variables recolectadas por la encuesta de ciberacoso; esta incluye dos partes, la primera fue asociada a variables demográficas y la segunda parte fueron las variables propias de la encuesta.

Entre las variables demográficas se incluyen aspectos como género, edad, lugar de nacimiento, estado civil, tipo de colegio, lugar de trabajo, entre otros. Mientras que las variables propias de la encuesta recolectaron valores basados en una escala de Likert, para medir el ciberacoso a través del uso de dispositivos electrónicos más comunes.

Los datos recolectados, que se usaron en esta investigación, corresponden a las encuestas realizadas en los meses de mayo y junio del 2022, a estudiantes entre 18 y 22 años de edad de una universidad. Cada pregunta realizada se calificó dentro de una escala ordinal para obtener la frecuencia de victimización, los actores del acoso y patrones de comportamiento que puedan presentarse en la encuesta (Useche et al., 2021).

Análisis de los datos y selección de características

Dentro de esta fase fue necesaria la limpieza y transformación de datos, debido a que los datos existentes se presentaron en diferentes formatos, incoherencias, así como datos faltantes o incorrectos.

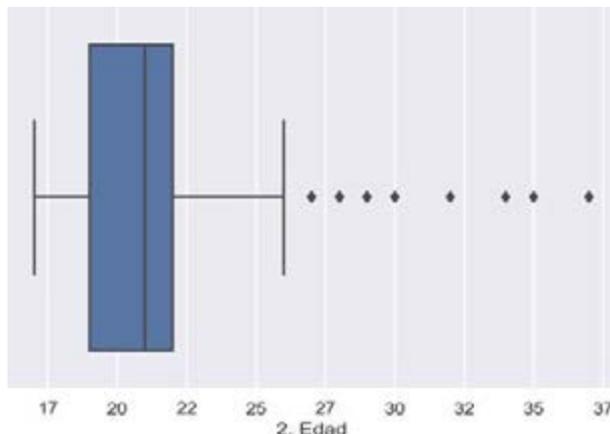
Los datos recolectados en la encuesta corresponden a un grupo de adolescentes con una edad entre los 18 y 22 años, con un nivel socioeconómico medio alto, que asisten a una universidad de la ciudad de Cuenca, Ecuador. Se aplicó la encuesta considerando una población de 7,000 estudiantes con un nivel de confianza del 99% y un margen de error del 5%. Con estos parámetros, el tamaño de la muestra resultante es de 607 encuestas tomadas, que se dividen en 286 mujeres y 316 hombres, siendo solo en 5 encuestas donde prefirieron no responder. Dentro de esta muestra resultante no se encontraron valores faltantes, por lo que se pudo considerar como una muestra completa.

Análisis exploratorio

En este análisis de los datos se observaron las variables correspondientes a los resultados de la encuesta. Como se mencionó, de las 607 encuestas realizadas, cada una de ellas contiene 52 variables y se divide en dos secciones. La primera sección contiene variables de información sociodemográfica de los alumnos y preguntas de encuesta en escala sobre el ciberacoso, referentes a la victimización y la agresión. En cambio, en la segunda sección de la encuesta, las respuestas estaban relacionadas con una escala de Likert que consta de cuatro valores cuyos identificadores son “nunca”, “algunas veces”, “bastantes veces”, “muchas veces”.

Dentro de las variables de la encuesta existe una variable numérica correspondiente a la edad del encuestado; se analizó su distribución y los valores atípicos que pueda presentar. Como se observa en la Figura 2, la variable edad sigue una distribución normal y muestra valores atípicos que han sido tratados por un proceso de imputación de reemplazo por el límite superior.

Figura 2
Distribución de la edad de los encuestados



Sin embargo, el resto de las variables fueron definidas como categóricas, determinadas por el porcentaje de cada categoría. Por lo tanto, aquellas categorías que no llegaron a un 5% de los valores fueron reasignadas a otras categorías, siendo un proceso lógico de agrupación. Como se observa, la Tabla 1 presenta un ejemplo de una variable categórica en conjunto con la cantidad de respuestas y su correspondiente porcentaje.

Tabla 1
Ejemplo de la variable categórica con su respuesta y porcentaje

VARIABLE	CATEGORIA	RESPUESTAS	PORCENTAJE (%)
"He amenazado a alguien para meterle miedo"	Nunca	552	90.93
	Algunas veces	45	7.41
	Bastantes veces	3	0.49
	Muchas veces	3	0.49
	No contesta	4	0.65

De esta manera, se puede observar que en la Tabla 1, las categorías "Bastantes veces", "Muchas veces" y "No contesta" no llegan a superar el 1% de las respuestas del grupo entrevistado. Por lo tanto, estas categorías fueron agrupadas para el proceso de generación del clúster. Este proceso fue repetido en todas las variables categóricas de la encuesta.

Selección de variables

Para la selección de las variables se utilizó el algoritmo de Stepwise (Görzig & Frumkin, 2013) sobre un proceso de regresión logística para brindar una relación entre una variable dependiente respecto de otras variables independientes. Por lo tanto, en este algoritmo se realiza una construcción iterativa, paso a paso, de un modelo de regresión, que implica la selección de variables independientes, que se utilizan en un modelo final. En este estudio se usó la regresión logística ya que se requiere determinar si una persona sufrió o no ciberacoso; en otras palabras, las variables de interés para este estudio son dicotómicas (toman dos valores: sí o no). De este

modo, la aplicación del modelo de regresión logística permitió construir una función que se basa en el cálculo de la probabilidad donde la variable de interés adopta el valor del evento previamente definido a través de la ecuación $y = \ln(p/(1-p))$, donde p es la probabilidad que ocurra un evento Y , que implica a los coeficientes de los términos constantes y variables, dependientes e independientes.

Así, la nueva variable dependiente generada, que se va a estimar, puede llegar a tener cualquier valor dentro de un rango. En este sentido, al aplicar el modelo de regresión logística implica agregar o eliminar variables explicativas potenciales en sucesión y probar la significancia estadística después de cada iteración (Sultan et al., 2021). Por ende, se utilizaron todas las variables numéricas y categóricas en el proceso de selección de variables, y además, la variable de interés para determinar la significancia final es “*Personas que han recibido acoso*”. En la Tabla 2 se presentan las variables con mayor valor significativo.

Tabla 2
Selección de variables y sus valores de significancia

VARIABLE	ESTIMACIÓN	Z	PROBABILIDAD
“Ciclo”	0.10390	1.664	0.09614
“Celular insultado ridiculizado”	27.85917	0.012	0.99043
“Celular obligado a hacer cosas que no quería dinero hacer tareas”	17.58782	0.004	0.99645
“Celular contado mentiras rumores sobre mí”	1.18124	2.407	0.01607
“Celular han pasado manipulado fotos videos de mí sin mi permiso”	2.20479	2.837	0.00455
“Internet me han insultado o ridiculizado”	1.33325	1.91	0.05612
“Internet han contado mentiras o rumores sobre mí”	0.29532	0.962	0.33594
“Móvil internet he obligado a alguien a hacer cosas con amenazas”	19.37827	0.009	0.99283
“Móvil internet he mentido en el Messenger o en cuentas privadas de otros”	1.20013	0.937	0.34889
“Cree alguna vez haber sufrido burlas o acoso en su lugar de estudio”	1.82379	7.036	1.98E-12

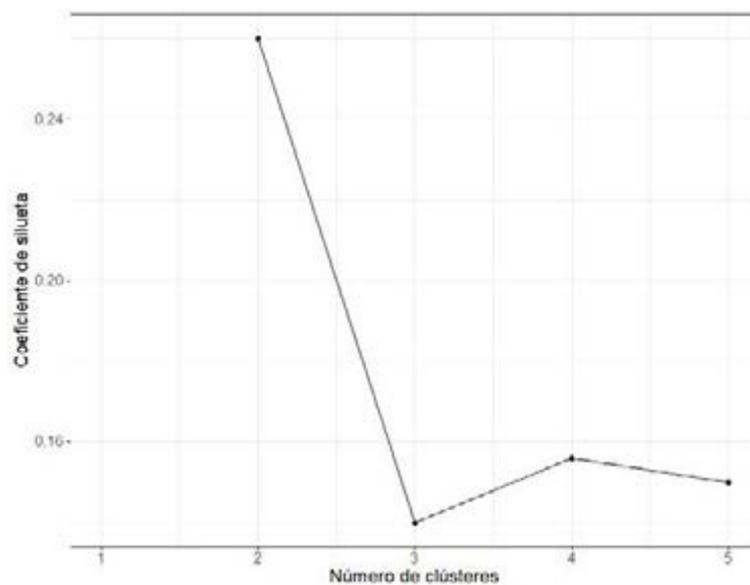
Definición y aplicación del modelo

En esta fase, se aplicó la técnica de agrupamiento para identificar el patrón de comportamiento de quienes cometen acoso, detectando los grupos que pueden variar según diversos criterios demográficos. Una vez que se tuvieron las variables seleccionadas, se procedió a la aplicación de la clusterización de los elementos que contienen las clases de objetos similares. Esta técnica descriptiva intenta reconocer patrones similares dentro de un conjunto de datos, llegando a ser un algoritmo de aprendizaje no supervisado (Carballido et al., 2018). Sin embargo, todas las variables utilizadas son de tipo categórico, por lo que se debe utilizar un proceso de cálculo de distancias que permita calcular estos valores para las variables categóricas y luego puedan ser usadas en un algoritmo adecuado para variables categóricas.

Pues así, en este estudio se utilizó la distancia de Gower que puede medir las diferencias de dos registros distintos (Meaney & Fikes, 2023). Estos registros pueden presentarse como una combinación de datos lógicos, categóricos, numéricos o texto. Para obtener la distancia de Gower se realiza un promedio de las diferencias parciales entre los individuos.

Después de obtener las distancias, se procedió a determinar cuál es el número de clústeres óptimos para los individuos que realizaron la encuesta. Para ello, se utilizó una métrica que calcula la calidad del agrupamiento realizado con algoritmos de clusterización denominada coeficiente de silueta. Esta métrica oscila sus valores entre 1 siendo el mejor agrupamiento y -1 representando un mal agrupamiento. Además, los valores que llegan a ser cercanos a 0, indican que el agrupamiento está superpuesto o indiferente. De esta manera, el primer resultado para la selección de variables demostró que el número óptimo de agrupamiento para el grupo encuestado es dos clústeres, como se puede apreciar en la Figura 3.

Figura 3
Número óptimo de clústeres

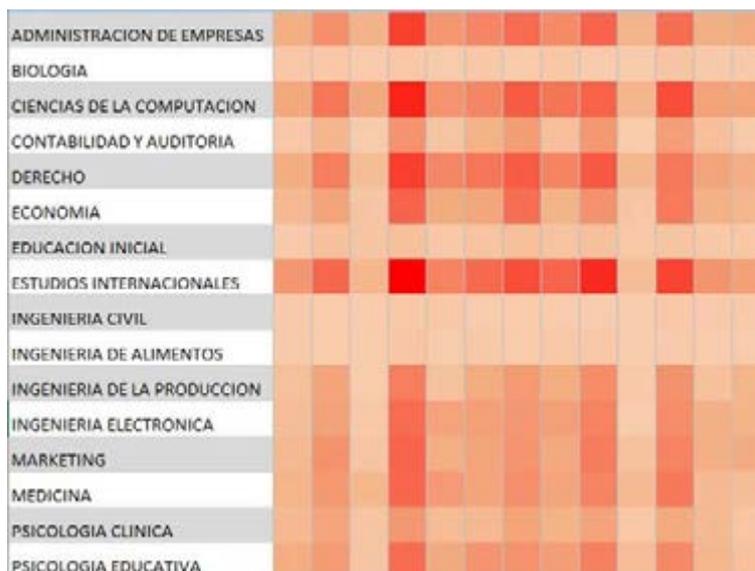


Una vez que se estableció el número de clústeres, se aplicó el algoritmo de asignación del clúster para cada uno de los individuos. Se empleó la distancia de Gower, para encontrar las medidas entre los sujetos estudiados; y se utilizó el algoritmo PAM, para la generación de los clústeres. Este algoritmo encuentra una secuencia de objetos denominados *medoides* que están localizados en la parte central del clúster (Schubert & Rousseeuw, 2021) y con base en estos *medoides*, los individuos son asignados a cada uno de los clústeres definidos.

Resultados

En esta parte se validó el análisis de los clústeres realizados para luego evaluar los resultados obtenidos. Para lo cual, se elaboraron varios modelos que poseen la suficiente información desde una perspectiva de análisis de datos, de tal modo que se puedan analizar y verificar los resultados de manera efectiva en relación con las variables demográficas. Por lo tanto, se asignaron los clústeres a cada uno de los individuos de la encuesta, y se pudo observar que la carrera de Estudios Internacionales, que se imparte en la universidad de estudio, presentó un alto índice de acoso y ciberacoso frente a otras carreras. Esto se puede observar en la Figura 4, que ofrece una perspectiva de un incremento de los niveles de acoso con una tendencia hacia el rojo.

Figura 4
Nivel de acoso y ciberacoso por carrera



Del mismo modo, se puede visualizar por ciclo el índice de acoso y ciberacoso con una marcada tendencia hacia el rojo, como se observa en la Figura 5. Así, el sexto ciclo fue donde se presentó un mayor índice de acoso, seguido del cuarto y segundo ciclo, pero el octavo ciclo tiene un menor grado de acoso.

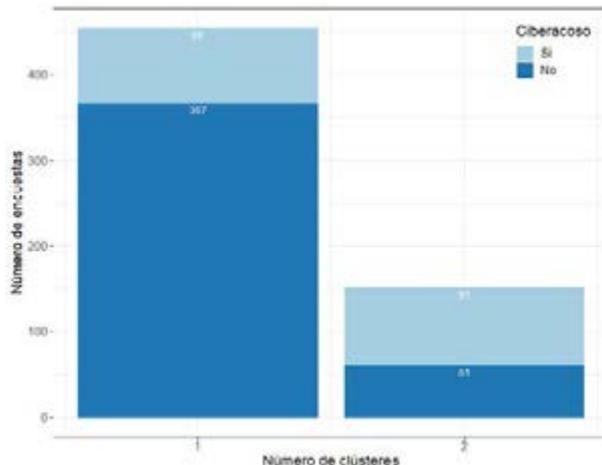
Figura 5
Nivel de acoso y ciberacoso por ciclo académico



Continuando con el análisis de los resultados, lo primordial fue determinar si la variable de interés está claramente identificada en algún clúster. Por lo tanto, las personas que recibieron acoso o ciberacoso fueron representadas con un valor de uno, y a aquellos estudiantes que estuvieron absueltos del acoso o ciberacoso se los representó con un cero, como se puede observar en la Figura 6.

Siendo así, un 29.49% de la cantidad de estudiantes encuestados habían recibido ciberacoso, por lo que se pudo identificar las características que se presentaron en este grupo de estudiantes. Dado que la variable de interés identifica claramente dos características dicotómicas (sí/no), los dos clústeres obtenidos fueron suficientes para obtener una conclusión relevante para el estudio.

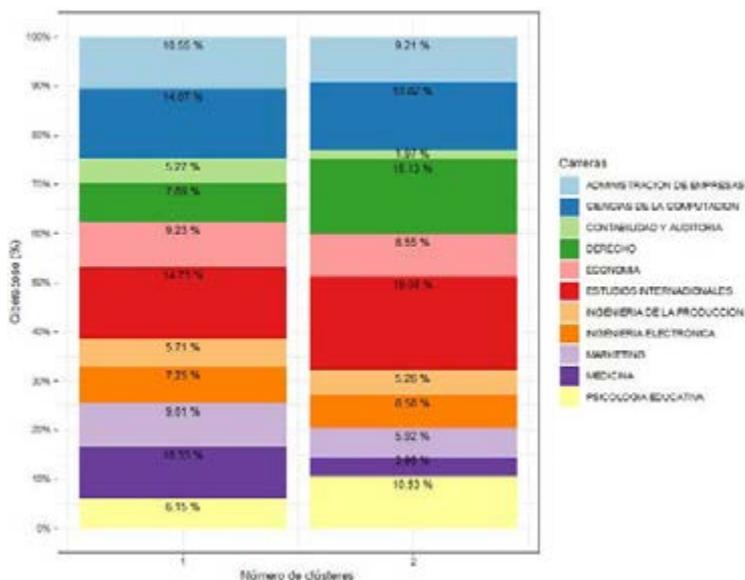
Figura 6
Cantidad de personas que recibieron ciberacoso



Sin embargo, para brindar un análisis con mayor detalle, fue necesario emplear cada una de las variables utilizadas en el modelo con la designación del clúster para los individuos del estudio. Para ello, se enfocó en las características de la agrupación de los dos clústeres utilizando gráficas porcentuales para su análisis respectivo.

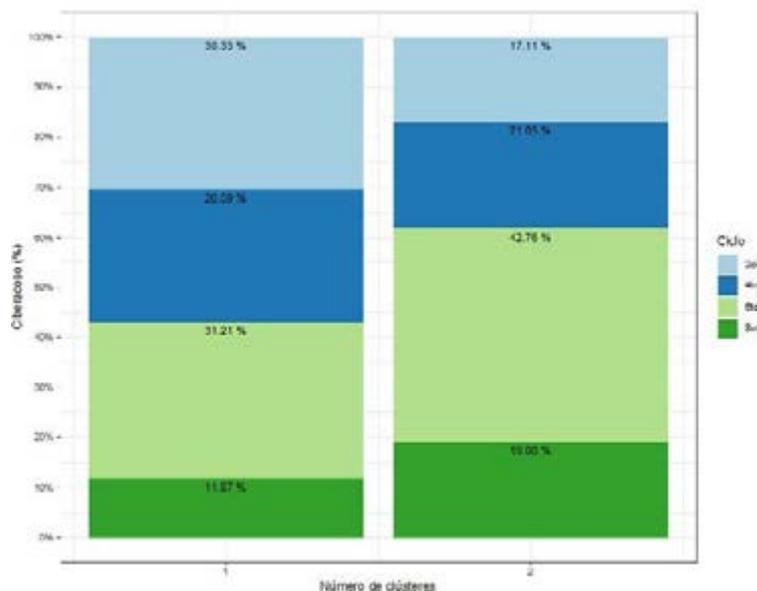
Una de las primeras variables que se analizó fue las carreras de la universidad que presentaron un mayor índice de ciberacoso en base a las características de la agrupación de los clústeres. Por lo tanto, existe una mayoría relativa en ambos clústeres con las carreras de Estudios Internacionales, Derecho y Ciencias de la Computación como se puede observar en la Figura 7.

Figura 7
Porcentaje de ciberacoso por carreras



Debido a la existencia de una mayoría relativa del ciberacoso en las carreras, como se presenta en la Figura 7, fue necesario conocer en cuáles ciclos lectivos se presenta una mayor incidencia del ciberacoso en la universidad de estudio. La tendencia que se presentó en el análisis de esta variable fue que los estudiantes del sexto ciclo presenciaron más ciberacoso que el resto de ciclos lectivos estudiados, como se puede observar en la Figura 8.

Figura 8
Porcentaje de ciberacoso por ciclos lectivos



Conclusiones

El presente estudio sobre la detección del ciberacoso en los estudiantes de una universidad de la ciudad de Cuenca, Ecuador, a través de la clusterización de los datos provenientes de las encuestas psicológicas, evidencia un claro panorama de la existencia de ciberacoso como un problema social y psicológico para los estudiantes. Gracias a los avances en inteligencia artificial y en minería de datos, se puede detectar e identificar el ciberacoso de forma acertada para afrontar esta problemática.

Por esta razón, al aplicar la minería de datos se obtienen resultados que son evidentes y se ajustan a la realidad de la problemática. Siendo así, este estudio fue analizado con el algoritmo de agrupamiento, separando cada una de las categorías de la encuesta y se obtuvieron comportamientos de los encuestados en cada pregunta analizada. Por lo tanto, los resultados demuestran claramente que la carrera de Estudios Internacionales es donde se presenta una mayor tasa de estudiantes implicados en el ciberacoso, con un 25.23%. Asimismo, en el sexto ciclo lectivo se detectó que un 42.86% de los estudiantes reciben ciberacoso en la universidad de estudio.

En trabajos futuros, se puede analizar el ciberacoso que reciben los estudiantes, brindando más información a la clusterización para identificar características particulares de esta problemática social. De igual forma, se puede realizar el mismo procedimiento para trabajar en otras problemáticas sociales y psicológicas como la violencia contra la mujer, y llegar a establecer las posibles víctimas de femicidio e incluso evitarlo de antemano.

Reconocimientos

Este trabajo ha sido financiado parcialmente por el Vicerrectorado de Investigaciones de la Universidad del Azuay, y forma parte de la Tesis titulada Aplicación de técnicas de clusterización en datos provenientes de la aplicación de una encuesta de cyberbullying. Los autores desean expresar su agradecimiento al Laboratorio de Investigación y Desarrollo en Informática por su participación en el trabajo.

Referencias

- Botyarov, M., & Miller, E. E. (2022). Partitioning around medoids as a systematic approach to generative design solution space reduction. *Results in Engineering*, 15, 100544. <https://doi.org/10.1016/j.rineng.2022.100544>
- Carballido, J. A., Latini, M. A., Ponzoni, I., & Cecchini, R. L. (2018). An Evolutionary Algorithm for Automatic Recommendation of Clustering Methods and its Parameters. *Electronic Notes in Discrete Mathematics*, 69, 229-236. <https://doi.org/10.1016/j.endm.2018.07.030>
- Dedic, G., Djordjevic, B., & Dedic, S. (2019). Victimization in childhood as a suicide risk factor in adults. *Vojnosanitetski pregled*, 76(7), 667-674. <https://doi.org/10.2298/VSP170826142D>
- Farag, N., El-Seoud, S. A., McKee, G., & Hassan, G. (2019). Bullying Hurts. *Proceedings of the 2019 8th International Conference on Software and Information Engineering*, 85-90. <https://doi.org/10.1145/3328833.3328869>
- Garmendia Larrañaga, M., Jiménez Iglesias, E., & Larrañaga Aizpuru, N. (2019). Bullying and cyberbullying: victimisation, harassment, and harm. The need to intervene in the educational centre. *Revista Española de Pedagogía*, 77(273). <https://doi.org/10.22550/REP77-2-2019-08>
- Görzig, A., & Frumkin, L. A. (2013). Cyberbullying experiences on-the-go: When social media can become distressing. *Cyberpsychology: Journal of Psychosocial Research on Cyberspace*, 7(1). <https://doi.org/10.5817/CP2013-1-4>
- Hosseinmardi, H., Ghasemianlangroodi, A., Han, R., Lv, Q., & Mishra, S. (2014). Towards understanding cyberbullying behavior in a semi-anonymous social network. *2014 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2014)*, 244-252. <https://doi.org/10.1109/ASONAM.2014.6921591>
- Johnson, P. (2022). *What makes a cyber bully/victim?: factors associated with the perpetration of cyberbullying by cybervictims*. Macquarie University.
- Meaney, M., & Fikes, T. (2023). The Promise of MOOCs Revisited? Demographics of Learners Preparing for University. *Journal of Learning Analytics*, 10(1), 113-132. <https://doi.org/10.18608/jla.2023.7807>
- Meliana, N., Sunardi, & Fadlil, A. (2019). Identification of Cyber Bullying by using Clustering Methods on Social Media Twitter. *Journal of Physics: Conference Series*, 1373(1), 012040. <https://doi.org/10.1088/1742-6596/1373/1/012040>
- Moya-Solís, A., & Moreta-Herrera, R. (2022). Víctimas de cyberbullying y su influencia en las Dificultades de Regulación Emocional en adolescentes del Ecuador. *Psychology, Society & Education*, 14(1), 67-75. <https://doi.org/10.21071/psye.v14i1.14066>
- Pamuji, A., & Setiawan, H. S. (2022). Cyberbullying Prediction As Cyber Counseling Tools With Data Mining Classification. *Bit (Fakultas Teknologi Informasi Universitas Budi Luhur)*, 19(1), 29. <https://doi.org/10.36080/bit.v19i1.1789>
- Parime, S., & Suri, V. (2014). Cyberbullying detection and prevention: Data mining and psychological perspective. *2014 International Conference on Circuits, Power and Computing Technologies [ICCPCT-2014]*, 1541-1547. <https://doi.org/10.1109/ICCPCT.2014.7054943>
- Schubert, E., & Rousseeuw, P. J. (2021). Fast and eager k-medoids clustering: O(k) runtime improvement of the PAM, CLARA, and CLARANS algorithms. *Information Systems*, 101, 101804. <https://doi.org/10.1016/j.is.2021.101804>
- Sitnik-Warchulska, K., Wajda, Z., Wojciechowski, B., & Izydorczyk, B. (2021). The Risk of Bullying and Probability of Help-Seeking Behaviors in School Children: A Bayesian Network Analysis. *Frontiers in Psychiatry*, 12. <https://doi.org/10.3389/fpsy.2021.640927>

- Sultan, D., Suliman, A., Toktarova, A., Omarov, B., Mamikov, S., & Beissenova, G. (2021). Cyberbullying Detection and Prevention: Data Mining in Social Media. *2021 11th International Conference on Cloud Computing, Data Science & Engineering (Confluence)*, 338-342. <https://doi.org/10.1109/Confluence51648.2021.9377077>
- Useche, S. A., Valle, E., Valle-Escolano, R., & Colomer-Pérez, N. (2021). Psychometric properties, validity and insights of the School Bullying Questionnaire (CIE-A) in secondary schools of the Valencian Community (Spain). *PLOS ONE*, *16*(11), e0259392. <https://doi.org/10.1371/journal.pone.0259392>
- Zhang, S. (2021). From Flaming to Incited Crime: Recognising Cyberbullying on Chinese WeChat Account. *International Journal for the Semiotics of Law - Revue internationale de Sémiotique juridique*, *34*(4), 1093-1116. <https://doi.org/10.1007/s11196-020-09790-x>

Algoritmo para Clasificar la Resolución de Conflictos en Aspirantes Automotrices

Algorithm to Classify Conflict Resolution among Automotive Applicants

Carlos Alberto Espinosa-Pinos¹ <https://orcid.org/0000-0002-7841-8090>, Juan Carlos Villota-Zambrano¹ <https://orcid.org/0009-0004-3190-3912>, Héctor Alberto Luzuriaga-Jaramillo² <https://orcid.org/0000-0002-7844-1805>

¹Universidad Indoamérica, Ambato, Ecuador
carlospinos@indoamerica.edu.ec,
jvillota3@indoamerica.edu.ec

²Universidad Técnica de Ambato, Ambato, Ecuador
ha.uzuriaga@uta.edu.ec



Esta obra está bajo una licencia internacional
Creative Commons Atribución-NoComercial 4.0

Enviado: 2023/07/16

Aceptado: 2023/08/21

Publicado: 2023/10/15

Resumen

La inadecuada capacidad para resolver conflictos en estudiantes de ingeniería automotriz puede tener consecuencias negativas en el ámbito laboral. El desarrollo del pensamiento lógico matemático puede ayudar a los estudiantes a desarrollar habilidades de análisis crítico, mejora de la capacidad de resolución de problemas, desarrollo de habilidades de razonamiento y comunicación efectiva, lo que les permite abordar los conflictos de manera efectiva y encontrar soluciones creativas. Esta investigación pretende identificar predictores de la capacidad para resolver problemas usando algoritmos de clasificación. Metodología: En este estudio se aplicaron tres algoritmos de clasificación y se usó el proceso KDD para identificar predictores de la capacidad para resolver problemas. El conjunto de datos cuenta con 60 registros de estudiantes de la carrera de ingeniería automotriz de la Universidad Equinoccial en Quito, Ecuador, a los que se les aplicaron tres herramientas: una ficha sociodemográfica, un test de Shatnawi relacionado con el pensamiento lógico matemático y un test de Watson Glaser sobre capacidad de resolución de conflictos. Resultados: El mejor modelo de clasificación es el algoritmo K-vecinos más cercanos y su capacidad de predicción es muy buena, con una tasa de verdaderos positivos

Sumario: Introducción, Materiales y Métodos, Resultados, Discusión de Resultados y Conclusiones.

Como citar: Espinosa, C., Villota, J. & Luzuriaga, H. (2023). Algoritmo para Clasificar la Resolución de Conflictos en Aspirantes Automotrices. *Revista Tecnológica - Espol*, 35(2), 170-180. Recuperado a partir de <http://www.rte.espol.edu.ec/index.php/tecnologica/article/view/1064>

frente a la tasa de falsos positivos AUC de 0.75, junto a un buen desempeño en clasificar casos negativos. El modelo puede ser mejorado con un aumento de la muestra, validación cruzada o ajuste de hiper parámetros. Conclusión: La edad y el pensamiento lógico matemático están fuertemente asociados con la capacidad de resolver conflictos. En futuras investigaciones es importante considerar variables adicionales como la experiencia en proyectos de resolución de problemas, el conocimiento técnico y la habilidad comunicativa; explorar el uso de algoritmos de aprendizaje automático más avanzados; diseñar intervenciones educativas específicas basadas en el desarrollo del pensamiento lógico matemático; o comparar la capacidad de resolución de conflictos entre diferentes disciplinas de ingeniería.

Palabras clave: Algoritmos de clasificación, aprendizaje automático, pensamiento lógico matemático, psicología, resolución de conflictos.

Abstract

Inadequate conflict resolution skills among automotive engineering students can negatively affect the workplace. Mathematical and logical thinking can help students develop critical analysis skills, problem-solving ability, reasoning skills, and effective communication, enabling them to deal effectively with conflicts and find creative solutions. This research aims to identify predictors of problem-solving ability using sorting algorithms. Methodology: In this study, three classification algorithms were applied, and the KDD process was used to identify predictors of problem-solving ability. The data set includes 60 records of students from the automotive engineering program at the Universidad Equinoccial in Quito, Ecuador, to whom three tools were applied: a sociodemographic card, a Shatnawi test related to mathematical logical thinking, and a Watson Glaser test on conflict resolution ability. Results: The best classification model is the K-nearest neighbor's algorithm; its predictive ability is excellent, with an actual positive rate versus a false positive rate AUC of 0.75 and a good performance in classifying negative cases. The model can be improved with an increased sample size, cross-validation, or hyper-parameter adjustment. Conclusion: Age and mathematical and logical thinking are strongly associated with conflict resolution ability. Future research: It is essential to consider additional variables such as experience in problem-solving projects, technical knowledge, and communication skills; to explore the use of more advanced machine learning algorithms; to design specific educational interventions based on the development of mathematical, logical thinking; or to compare conflict resolution ability between different engineering disciplines.

Keywords: Classification algorithms, machine learning, mathematical logical thinking, psychology, conflict resolution.

Introducción

En el campo de la ingeniería automotriz es común que los proyectos se realicen en equipos multidisciplinarios donde los estudiantes deben colaborar. Y comunicarse eficientemente. Los conflictos pueden surgir debido a diferencias en opiniones, enfoques o ideas. Rivera et al. (2023) afirman que la capacidad de resolver conflictos permite a los estudiantes abordar estas discrepancias de manera constructiva, encontrar soluciones y mantener un ambiente de trabajo armonioso. Los conflictos no resueltos pueden tener un impacto negativo en la productividad de un equipo. Si los estudiantes no saben cómo manejar adecuadamente los desacuerdos y conflictos, el progreso del proyecto puede retrasarse, lo que afecta el tiempo de entrega y la calidad del trabajo final.

Andajani-Sutjahjo et al. (2018) proclaman que la capacidad de resolución de conflictos permite a los estudiantes abordar rápidamente los problemas y mantener el flujo de trabajo sin interrupciones innecesarias. La resolución de conflictos eficaz fomenta un ambiente propicio para la innovación y la creatividad, puesto que cuando los estudiantes pueden expresar y discutir abiertamente sus ideas sin temor a conflictos negativos, se promueve la generación de nuevas soluciones y enfoques de innovación. La capacidad de resolver conflictos fomenta el pensamiento crítico y la búsqueda de alternativas, lo que puede llevar a avances significativos en el campo de la ingeniería automotriz.

La resolución de conflictos ayuda a construir y mantener relaciones interpersonales sólidas entre los estudiantes de ingeniería automotriz. A medida que trabajan juntos para superar los desafíos y conflictos, desarrollan habilidades de comunicación, empatía y comprensión mutua. Estas habilidades son valiosas tanto en el entorno académico como en el profesional, ya que facilitan la colaboración y el trabajo en equipo, según Celi Rojas et al. (2021). La capacidad de resolución de conflictos es importante para los estudiantes de ingeniería automotriz porque les permite trabajar de manera efectiva en equipo, mejorar la productividad, fomentar la innovación y mantener relaciones interpersonales sólidas. Estas habilidades son esenciales para el éxito en el campo de la ingeniería automotriz, donde el trabajo colaborativo y la resolución de problemas son fundamentales.

La ingeniería automotriz implica el uso intensivo de conceptos matemáticos, desde cálculo diferencial e integral hasta álgebra lineal y estadística. Un pensamiento lógico, matemático, sólido es fundamental para comprender y aplicar estos conceptos de manera efectiva (Alcívar Martínez, 2018). Medir el pensamiento lógico matemático en los estudiantes ayuda a identificar si tienen una base sólida en matemáticas según Rosero Ojeda (2021) y si están preparados para abordar los desafíos técnicos que se encontrarán en su campo de estudio según Núñez (2021). Los ingenieros automotrices se enfrentan a problemas complejos que requieren un enfoque lógico y analítico. La capacidad de razonar de manera lógica y aplicar conceptos matemáticos adecuados es esencial para descomponer problemas complicados en pasos más manejables y encontrar soluciones eficientes. Morales et al. (2018) describen que, al evaluar el pensamiento lógico matemático de los estudiantes, se puede determinar su capacidad para abordar problemas complejos y resolverlos de manera efectiva.

En el campo de la ingeniería automotriz, es necesario tomar decisiones informadas basadas en datos y análisis matemáticos. Esto puede incluir decisiones relacionadas con el diseño de vehículos, la optimización del rendimiento, el análisis de fallas, entre otros aspectos. Así mismo, Godoy Cedeño (2020) refiere que el pensamiento lógico matemático permite a los estudiantes evaluar y analizar datos de manera crítica, lo que les ayuda a tomar decisiones más fundamentadas y efectivas en su práctica profesional.

La ingeniería automotriz está en constante evolución y requiere la capacidad de pensar de manera creativa e innovadora. El pensamiento lógico matemático es una habilidad clave para plantear y resolver problemas complejos; según Yasig Salguero (2021), permite explorar nuevas ideas y desarrollar tecnologías innovadoras en el campo. Medir el pensamiento lógico matemático en los estudiantes permite identificar su capacidad para abordar desafíos técnicos y contribuir a la innovación en la industria automotriz.

La determinación o medición del pensamiento lógico matemático en los estudiantes de ingeniería automotriz es importante para garantizar que tengan una base sólida en matemáticas, puedan abordar problemas complejos, tomar decisiones informadas y contribuir a la innovación

en su campo. Estas habilidades son esenciales para el éxito en la ingeniería automotriz y para enfrentar los desafíos técnicos en constante cambio en la industria.

Algunos de los resultados de la búsqueda analizan la importancia del pensamiento matemático y lógico en la ingeniería, y la capacidad de identificar, formular y resolver problemas de ingeniería complejos utilizando los principios de la ingeniería, la ciencia y las matemáticas (Kolajo, 2020). Se sugiere que en cuanto a calidad y monitoreo de trabajo se refiere, la importancia de la probabilidad y la estadística en la toma de decisiones y reducción de riesgos en las actividades industriales es primordial, según Ortiz Aguilar et al. (2021). Además, algunos resultados discuten la importancia del razonamiento lógico y las habilidades de resolución de problemas para los ingenieros. La evaluación de habilidades del pensamiento computacional para predecir el aprendizaje y retención de estudiantes, si bien no se enfoca directamente en la capacidad de resolución de conflictos, proporciona información sobre la evaluación de las habilidades cognitivas en estudiantes de ingeniería según González-Quiñones et al. (2022). “El razonamiento lógico como requisito funcional en ingeniería”: Este artículo analiza los conceptos de lógica, resolución de problemas y razonamiento lógico como requisitos funcionales en ingeniería. Si bien no menciona específicamente la capacidad de resolución de conflictos, enfatiza la importancia del pensamiento lógico en la ingeniería. Según Rojas-López & García-Peñalvo (2020), “el desarrollo del pensamiento lógico matemático en los estudiantes costarricenses de undécimo, año de colegios académicos diurnos y su nivel de logro en el aprendizaje de las matemáticas”. Este estudio tuvo como objetivo aportar evidencia empírica respecto a la relación entre el desarrollo del pensamiento lógico matemático en estudiantes costarricenses de grado 11° en escuelas académicas diurnas y su nivel de aprovechamiento en matemáticas, específicamente en las áreas algebraica y geométrica. El estudio encontró un bajo nivel de desarrollo del pensamiento lógico matemático, significativamente alejado de los parámetros aceptados internacionalmente. Las escalas, el razonamiento lógico y la demostración matemática fueron las más desafiantes. El nivel de logro en álgebra y geometría también fue claramente insuficiente (Vargas Salazar, 2017). Aunque los resultados de la búsqueda no proporcionan estudios específicos relacionados con la capacidad de resolución de conflictos en estudiantes de ingeniería automotriz usando los algoritmos de clasificación, estos estudios ofrecen información sobre la evaluación de las habilidades cognitivas y la importancia del pensamiento lógico en estudiantes de educación superior y de secundaria. Es posible que se necesite más investigación para encontrar estudios que aborden específicamente la capacidad de resolución de conflictos en el ámbito académico de la ingeniería automotriz.

El test de Watson Glaser está formado por varias dimensiones, una de ellas es la inferencia que es una conclusión que una persona puede extraer de ciertos hechos observados o supuestos; reconocimiento de suposiciones: Una suposición es algo que se presupone o se da por sentado; La Deducción: En esta prueba, cada ejercicio consta de varias afirmaciones (premisas) seguidas de varias propuestas de conclusiones; Interpretación: Cada uno de los ejercicios propuestos consiste en un breve párrafo seguido de varias conclusiones sugeridas. Se supone que todo lo que aparece en el párrafo corto es cierto, pero el problema consiste en juzgar si cada una de las conclusiones propuestas se deduce lógicamente, más allá de toda duda razonable, de la información contenida en el párrafo; Evaluación de argumentos: A la hora de tomar decisiones sobre cuestiones importantes, es conveniente poder distinguir entre argumentos sólidos y argumentos débiles en lo que respecta a la cuestión. Para que un argumento sea sólido, debe ser importante y estar directamente relacionado con la cuestión. Un argumento es débil si no está directamente relacionado con la cuestión (aunque pueda tener una gran importancia general) (Possin, 2014).

La evaluación del pensamiento crítico de Watson Glaser. Es una prueba diseñada para evaluar la capacidad de un individuo para pensar críticamente. La prueba proporciona una puntuación total que se puede utilizar para clasificar a las personas según sus habilidades de pensamiento crítico. Un puntaje de al menos 70% en la prueba. Indica una capacidad avanzada de pensamiento crítico; un puntaje de al menos el 40% Se relaciona con un nivel básico de pensamiento crítico.; mientras que un puntaje o porcentaje menor al 40% indica problemas o un nivel deficiente en la habilidad de pensamiento crítico según Rivera et al. (2023). Una puntuación más alta en la prueba puede indicar habilidades de pensamiento crítico más sólidas que son esenciales para la resolución efectiva de conflictos.

Materiales y Métodos

Proceso de descubrimiento de conocimiento en base de datos. En este estudio se ha utilizado el proceso Knowledge Discovery Database (KDD), un proceso automático que combina descubrimiento y análisis para extraer patrones en forma de reglas en funciones de los datos. El proceso KDD consta de múltiples capas; su principal objetivo es extraer conocimientos valiosos, pero no triviales de los datos a los que tiene acceso. Las etapas del proceso KDD se describen a continuación: 1. Identificación del problema e identificación del dominio de trabajo. 2. Creación del conjunto de datos. 3. De procesamiento de los datos. 4. Reducción y proyección de los datos. 5. Formulación de los objetivos de la KDD. 6. Selección de métodos de minería de datos. 7. Evaluación de los patrones descubiertos. 8. Presentación y visualización. 9. Integración del conocimiento descubierto en el sistema (Camilo Giraldo Mejía & Alberto Vargas Agudelo, 2019).

Base de datos e instrumentos de recolección de datos

El conjunto de datos cuenta con 60 registros de estudiantes de la carrera de ingeniería automotriz de la Universidad tecnológica Equinoccial, a quienes se les aplicó 3 instrumentos: El primero corresponde a una ficha sociodemográfica; el segundo a una prueba para determinar el pensamiento lógico matemático denominado test de Shatnawi (Vargas Salazar, 2017); y el tercer instrumento corresponde a la prueba de Watson Glaser Relacionada con la medición del pensamiento crítico.

La variable a predecir se denomina BUENAS_CUALI la cual hace referencia a la cantidad de respuestas exitosas. Como resultado de la aplicación del test de Watson Glaser; es de tipo nominal y puede ser clasificada como deficiente o básica avanzada. Las variables predictoras son el sexo y el estado civil, ambas de tipo nominal; la edad, y las 6 dimensiones que forman parte del test que mide el desarrollo del pensamiento lógico matemático, estas últimas de tipo cuantitativas. La Tabla 1 muestra la descripción de las variables y sus posibles valores.

Tabla 1
Variables Analizadas

Variables	Código	Tipo
Género	SEXO	nominal
Estado civil	CIVIL	nominal
Edad	EDAD	cuantitativa
Generalización	GG	cuantitativa
Inducción	II	cuantitativa
Deducción	DD	cuantitativa

Variables	Código	Tipo
Uso de símbolos matemáticos	SS	cuantitativa
Razonamiento lógico	RR	cuantitativa
Demostración matemática	PP	cuantitativa
Global del test pensamiento lógico	DPLM_10	cuantitativa
Global del test Watson-Glaser	BUENAS_CUALI	cuantitativa

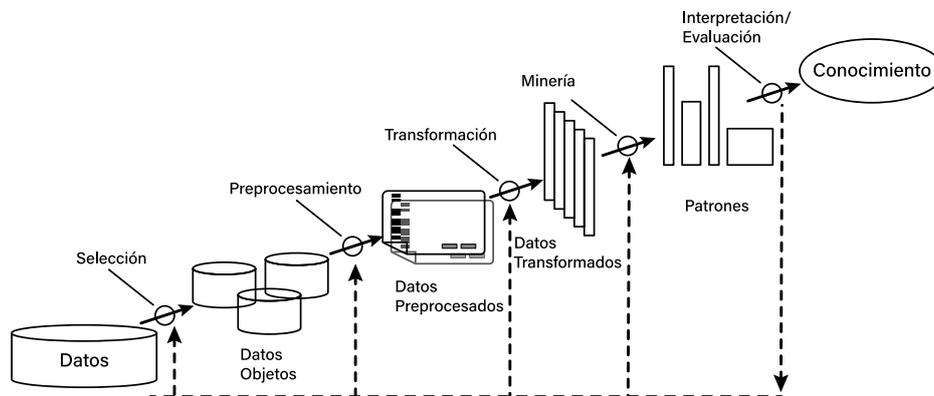
Descripción de variables

El género se refiere a cómo se auto identifican los alumnos estudiados, es decir si se identifican como hombres o mujeres para este caso específico. En el estado civil existen dos opciones que son: soltero o casado. La edad indica los años de vida de los estudiantes, que en promedio es de 22,5 años. Las dimensiones del test de desarrollo de pensamiento lógico matemático están conformadas por: La generalización, que comprende la búsqueda de modelos o modelización como un componente fundamental del pensamiento lógico matemático; la deducción, que se refiere al proceso que conlleva a conclusiones válidas, siempre y cuando las premisas de partida también lo sean. La inducción, que se caracteriza por conducir al hallazgo de modelos partiendo de casos específicos; el uso de símbolos y el uso del lenguaje matemático, que se refiere al proceso de generalización matemática, unificando su representación de forma concisa, precisa y clara. El razonamiento lógico, que se refiere a la habilidad para decidir si las proposiciones son o no lógicamente verdaderas; la capacidad de realizar demostraciones matemáticas, que se relaciona con los métodos usados por los estudiantes para validar sus construcciones teóricas o fundamentar los resultados obtenidos; el test global de pensamiento lógico, representado por DPLM_10, es el resultado global obtenido de aplicar el test de Shatnawi; y el test global de Watson-Glaser, llamado BUENAS_CUALI, que es el valor global obtenido al aplicar el test de pensamiento crítico (Aiyub et al., 2021).

Pasos para ejecutar el proceso KDD Se muestra en la Figura 1 en primer lugar los datos relativos a la valoración del pensamiento crítico de los estudiantes de ingeniería automotriz de la Universidad Tecnológica Equinoccial, de los últimos niveles séptimo y octavo, respectivamente son subidos en una hoja de Excel junto a los resultados obtenidos de la ficha sociodemográfica y el test de desarrollo del pensamiento lógico matemático aplicado a los 60 estudiantes. En este estudio se utilizó el desarrollo del pensamiento lógico matemático y su influencia en el pensamiento crítico. El segundo paso es la preparación de los datos, es decir almacenar los valores obtenidos en el pensamiento lógico matemático y la valoración global en un archivo de Excel en formato csv delimitado por comas para hacerlo visible en la aplicación estadística JASP. Es necesario indicar que en esta etapa se depura la base, verificando que los valores decimales están definidos por puntos y los nombres de las variables van sin tilde porque el idioma nativo del programa estadístico es el inglés; también se procede a la verificación del respectivo tipo de variable para finalmente verificar que no hay celdas vacías ni ningún símbolo especial, tanto en los datos como en los nombres de las variables según lo indicado en (Martin Guambugete Rea et al., 2023). En el tercer paso, los datos se modelizan mediante JASP para predecir el pensamiento crítico de los estudiantes de ingeniería automotriz a partir de las variables sociodemográficas, así como de las dimensiones y el global del test de desarrollo de pensamiento lógico matemático de Shatnawi; En este paso, hay una secuencia de 5 fases: entrenamiento, patrón de prueba, evaluación de resultados y conocimiento, representación.

Los datos depurados se dividen en dos partes: la fase de entrenamiento y la fase de prueba. Según Berástegui Arbeloa (2018), el algoritmo K-vecinos más cercanos para construir un modelo en la fase de entrenamiento, es un algoritmo que usa el método de decisión de tipo binario para clasificar el problema. Para evaluar el rendimiento del modelo de aprendizaje generado, utiliza la validación cruzada k-fold, para evaluar el rendimiento del modelo de aprendizaje generado. El algoritmo divide el conjunto de datos en K, subconjuntos o pliegues del mismo tamaño y, a continuación, el modelo se entrena k veces, cada una de ellas utilizando un subconjunto diferente como prueba y los otros K menos 1 subconjuntos como entrenamiento. Visualmente, la etapa de prueba se refleja en el área bajo la curva de características operativas del receptor (ROC), donde el clasificador seleccionado se muestra en función de las observaciones relacionadas con las categorías o dimensiones de la variable objetivo a predecir. Ver Figura 1.

Figura 1
Proceso KDD



Resultados y Discusión

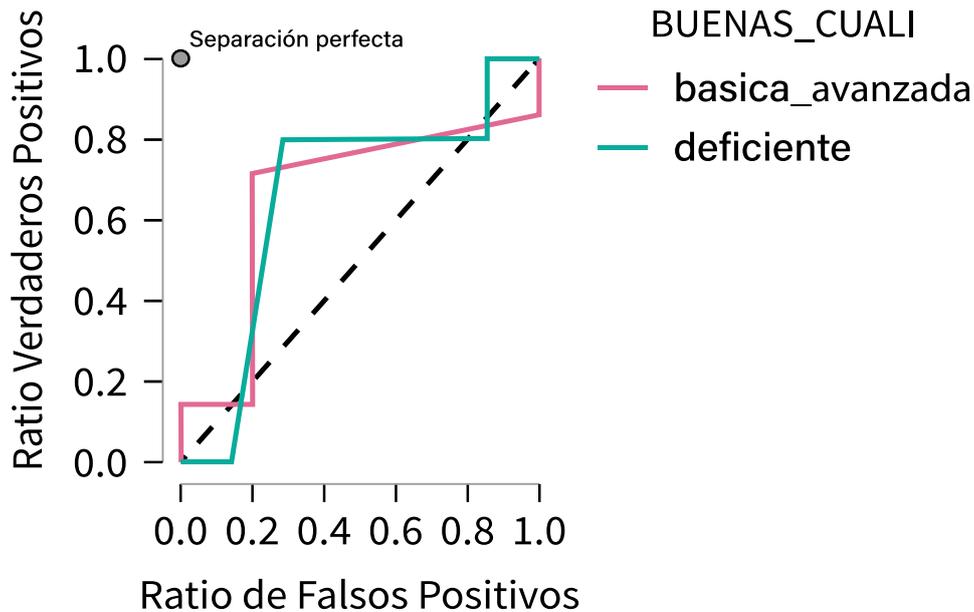
Se aplicaron tres algoritmos de clasificación tales como K-vecinos más cercanos, Bosque aleatorio, el árbol de decisión; de los cuales el mejor modelo de clasificación resultó ser el primero, es decir el modelo K-vecinos más cercanos. Esto se puede evidenciar en la siguiente Tabla 2, con las respectivas métricas de evaluación.

Tabla 2
Precisión de los Algoritmos de Clasificación

Clasificación de los K-Vecinos más cercanos							
Vecinos más próximos	Ponderaciones	Distancia	n (Entrenamiento)	n (Contraste)	Precisión del Contraste		
3	rank	Manhatan	48	12	0.7500		
Clasificación de Árbol de decisión							
Separadores	n (Entrenamiento)	n (Contraste)	Precisión del Contraste				
6	24	36	0.6111				
Clasificación en Bosques Aleatorios							
Árboles	Características por separador	n (Entrenamiento)	n (Validación)	n (Contraste)	Precisión de la Validación	Precisión del Contraste	Precisión de OOB
11	3	38	10	12	0.5000	0.5833	0.5789

La Figura 3, muestra el diagrama de curvas ROC, donde la evaluación de la capacidad del modelo para distinguir entre las diferentes categorías de la variable a predecir es muy buena. La tasa de verdaderos positivos frente a la tasa de falsos positivos (1- especificidad) bajo la curva AUC es de 0,75 lo que indica una alta capacidad discriminante. Aun así, podría ser mejor en clasificación, ya que es sensible al tamaño del conjunto de datos. Un aumento en la muestra de datos observados, así como técnicas de la validación cruzada o el ajuste de hiper parámetros. Para evaluar y mejorar el modelo actual, es una recomendación.

Figura 2
Diagrama de curvas ROC (JASP 16.0)



El valor predictivo negativo (VPN) es la proporción de casos negativos. Correctamente clasificadas en relación con todos los casos clasificados como negativos. En este caso, los valores son 0.6667, 0.8333 y 0.7500, respectivamente para las diferentes clases. Un VPN más alto indica que el modelo es más efectivo en predecir correctamente los casos negativos. Ratio de verdaderos negativos (RVN) es la proporción de casos negativos correctamente clasificados en relación con todos los casos reales negativos. Los valores que proporcionados son 0.8000, 0.7143 y 0.7571, respectivamente. Un RVN más alto indica que el modelo es más efectivo en identificar correctamente los casos que son realmente negativos. Ratio de falsos negativos (RFN) es la proporción de casos negativos que fueron clasificados incorrectamente como positivos en relación con todos los casos reales negativos. Un RFN más bajo indica que el modelo comete menos errores al clasificar incorrectamente los casos negativos como positivos. El ratio de falsos negativos RFN y el ratio de falsas comisiones RFO son relativamente bajos, lo que indica que el modelo comete menos errores al clasificar incorrectamente los casos negativos como positivos y viceversa. La puntuación de amenaza es relativamente baja, lo que sugiere que el modelo es efectivo en minimizar los errores de clasificación de falsos negativos en comparación con los verdaderos negativos. En general, estos indicadores muestran un buen desempeño del modelo K-vecinos más cercanos como clasificador. Ver Tabla 3.

Tabla 3
Métricas de evaluación (JASP 16.0)

Métricas de Evaluación			
	basica_avanzada	deficiente	Average / Total
Apoyo	7	5	12
Precisión	0.7500	0.7500	0.7500
Precisión (Valor Predictivo Positivo)	0.8333	0.6667	0.7639
Exhaustividad (Ratio de Verdaderos Positivos)	0.7143	0.8000	0.7500
Ratio de Falsos Positivos	0.2000	0.2857	0.2429
Ratio de Falsos Hallazgos	0.1667	0.3333	0.2500
Puntuación F1	0.7692	0.7273	0.7517
Coefficientes de Correlación de Matthews	0.5071	0.5071	0.5071
Área Bajo la Curva (ABC, AUC)	0.6571	0.6571	0.6571
Valor Predictivo Negativo	0.6667	0.8333	0.7500
Ratio de Verdaderos Negativos	0.8000	0.7143	0.7571
Ratio de Falsos Negativos	0.2857	0.2000	0.2429
Ratio de Falsas Omisiones	0.3333	0.1667	0.2500
Puntuación de Amenaza	1.2500	0.8000	1.0250
Paridad Estadística	0.5000	0.5000	1.0000

Nota. Todas las métricas se calculan para cada clase frente a todas las otras clases.

Conclusiones

De acuerdo al modelo de clasificación, K-vecinos más cercanos, las variables que mejor permiten clasificar al Pensamiento Crítico, son: Edad (EDAD) con un aporte individual al modelo de clasificación de 0.4167; Global de desarrollo de pensamiento lógico matemático (DPLM_10), Demostración matemática (PP), Razonamiento Lógico (RR), cada uno con un aporte individual de 0.25; luego, las dimensiones de Inducción, Deducción, y Uso de símbolos y Lenguaje matemático, con un aporte individual cada una de 0.1667; a continuación, la dimensión generalización (GG), con un aporte individual al modelo de 0.08333. Finalmente, la variable SEXO resta precisión al modelo, y la variable CIVIL tiene un comportamiento neutral, por lo cual las dos variables no son buenas predictoras. Este proceso se hizo manualmente, según Espinosa-Pinos et al. (2022), pues se retiraron las variables del modelo general clasificador, para determinar su impacto individual en el mismo.

Dado que la edad tiene un aporte significativo en la clasificación del pensamiento crítico, se pueden desarrollar estrategias y programas específicos adaptados a diferentes grupos de edad. Esto implica abordar de manera diferenciada las necesidades y características de los estudiantes más jóvenes y los más adultos, dándoles oportunidades de desarrollo y práctica acordes a su nivel de madurez cognitiva. La dimensión global del desarrollo del pensamiento lógico matemático ha demostrado ser un predictor importante del pensamiento crítico. Por lo tanto, se puede proponer implementar actividades y recursos educativos que promuevan el desarrollo de habilidades en esta área. Esto podría incluir ejercicios de razonamiento lógico matemático, problemas desafiantes que requieran de inducción y deducción, la exploración de símbolos y el lenguaje matemático en contextos reales y relevantes para la ingeniería automotriz. Dado que tanto la demostración matemática como el razonamiento lógico contribuyen significativamente a la clasificación del

pensamiento crítico, se pueden diseñar actividades específicas que promuevan y desarrollen estas habilidades. Esto implica brindar oportunidades para que los estudiantes practiquen la resolución de problemas usando demostraciones lógicas y razonamiento matemático riguroso. Se puede proponer diseñar intervenciones educativas basadas en los resultados y variables identificadas como relevantes en el modelo de clasificación. Luego, se debe evaluar el impacto de estas intervenciones en la mejora del pensamiento crítico de los estudiantes de ingeniería automotriz. Esto permitirá verificar la efectividad de las estrategias implementadas y ajustarlas según sea necesario.

En futuras investigaciones sobre la capacidad de resolución de conflictos en estudiantes de ingeniería automotriz, es importante considerar variables adicionales, la experiencia en proyectos de resolución de problemas, el conocimiento técnico y la habilidad de comunicación. Además, podría explorar el uso de algoritmos de aprendizaje automático más avanzados que mejoran la precisión de la clasificación. Un estudio longitudinal que analice el desarrollo de estas habilidades a lo largo del tiempo proporcionaría información valiosa sobre los factores que influyen en su mejora.

Referencias

- Aiyub, Suryadi, D., Fatimah, S., & Kusnandi. (2021). Investigation of Watson-Glaser critical thinking skills of junior high school students in solving mathematical problems. *Journal of Physics: Conference Series*, 1806(1), 012090. <https://doi.org/10.1088/1742-6596/1806/1/012090>
- Alcívar Martínez, M. Y. (2018). *Instrumento para el diagnóstico del aprendizaje de la Geometría en estudiantes de octavo grado de educación general básica superior de la Unidad Educativa "la Maná"*. <https://repositorio.uti.edu.ec/handle/123456789/982>
- Andajani-Sutjahjo, S., Liew, T. C. H., Smith, J. F., Esekielu, I., Mason, G., & Tariu, I. (2018). Engaging community volunteers in participatory action research in Tamaki community of Auckland, New Zealand. *Health Promotion International*, 33(2), 219–228. <https://doi.org/10.1093/HEAPRO/DAW057>
- Berástegui Arbeloa, G. (2018). *Implementación y comparación del algoritmo de los k vecinos más cercanos (k-NN) con valores locales en k*. <https://academica-e.unavarra.es/xmlui/handle/2454/29112>
- Camilo Giraldo Mejía, J., & Alberto Vargas Agudelo, F. (2019). Aplicación de la técnica regresión logística de la minería de datos en el proceso de descubrimiento de conocimiento (KDD) en bases de datos operativas o transaccionales. *Perspectiv@s*, 14(13), 51–55. <http://revistas.uigv.edu.pe/index.php/perspectiva/article/view/568>
- Celi Rojas, S. Z., Sánchez, V. C., Quilca Terán, M. S., Paladines Benítez, M. del C., Celi Rojas, S. Z., Sánchez, V. C., Quilca Terán, M. S., & Paladines Benítez, M. del C. (2021). Estrategias didácticas para el desarrollo del pensamiento lógico matemático en niños de educación inicial. *Horizontes Revista de Investigación En Ciencias de La Educación*, 5(19), 826–842. <https://doi.org/10.33996/REVISTAHORIZONTES.V5I19.240>
- Espinosa-Pinos, C. A., Ayala-Chauvín, I., & Buele, J. (2022). Predicting Academic Performance in Mathematics Using Machine Learning Algorithms. *Communications in Computer and Information Science*, 1658 CCIS, 15–29. https://doi.org/10.1007/978-3-031-19961-5_2/COVER
- Godoy Cedeño, C. E. (2020). Uso de la gamificación en el desarrollo del pensamiento lógico matemático en estudiantes de educación superior en una universidad privada de Lima, 2020. *Repositorio Institucional - UCV*.
- González-Quiñones, F., Tarango, J., Javier, & Bojorquez, K. (2022). Ansiedad matemática en estudiantes mexicanos de ingeniería: correlación con el pensamiento matemático y el razonamiento Covariacional. *Revista de Psicología y Ciencias Del Comportamiento de La Unidad Académica de Ciencias Jurídicas y Sociales*, 13(2), 68–89. <https://doi.org/10.29059/RPCC.20221201-151>

- Kolajo, Y. (2020). *Developing Critical Thinking in a First-Year University Chemistry Course*.
- Martin Guambuguete Rea, C. I., Elizabeth Castro Mayorga III, M., Alberto Espinosa Pinos, C. I., Augusta Sánchez Benítez, C. I., & Rodrigo Jordan Bolaños, C. V. (2023). Factores asociados al rendimiento académico: Un análisis mediante regresión logística multivariante en estudiantes del primer nivel del Instituto Superior Tecnológico Tres de Marzo de la Provincia Bolívar, Ecuador. *Domino de Las Ciencias*, 9(3), 570–589. <https://doi.org/10.23857/DC.V9I3.3460>
- Morales, G. P., Gavilanes, D. A., & Jurado, D. B. (2018). Desarrollo del pensamiento lógico-matemático a través de juegos populares y tradicionales en niños de educación inicial. *Revista Científica Ciencia y Tecnología*, 18(19), 133–141. <https://doi.org/10.47189/RCCT.V18I19.190>
- Ortiz Aguilar, W., Ortega Chávez, W., Valencia Cruzaty, L. E., González Vásquez, Á. E., Gamarra Mendoza, S., Ortiz Aguilar, W., Ortega Chávez, W., Valencia Cruzaty, L. E., González Vásquez, Á. E., & Gamarra Mendoza, S. (2021). La educación estadística del ingeniero: reto de la educación superior. *Revista Universidad y Sociedad*, 13(5), 307–318. http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S2218-36202021000500307&lng=es&nrm=iso&tlng=pt
- Posgrado, D. DE, & Núñez Elisa, 2021. (2021). *El aprendizaje adaptativo asincrónico para el desarrollo de operaciones aritméticas combinadas con números reales, en los estudiantes de 8vo año de EGB de la Unidad Educativa Yanahurco del cantón Mocha*. <https://repositorio.uti.edu.ec/handle/123456789/2370>
- Possin, K. (2014). Critique of the Watson-Glaser Critical Thinking Appraisal Test: The more you know, the lower your score. *Informal Logic*, 34(4), 393–416. <https://doi.org/10.22329/IL.V34I4.4141>
- Rivera, M.-J., Matute, R., Rivera, M.-J., & Matute, R. (2023). Resolución de conflictos para estudiantes de psicología: Una propuesta de Cultura de Paz. *Psicoperspectivas*, 22(1), 31–45. <https://doi.org/10.5027/PSICOPERSPECTIVAS-VOL22-ISSUE1-FULLTEXT-2769>
- Rojas-López, A., & García-Peñalvo, F. J. (2020). Assessment of computational thinking skills to predict student learning and retention in the subject programming computer in higher education. *Revista de Educación a Distancia*, 20(63). <https://doi.org/10.6018/RED.409991>
- Rosero Ojeda, T. del R. (2021). *Estrategias de Mentoring Educativo para mejorar el rendimiento académico de matemáticas en niños de séptimo año*. <https://repositorio.uti.edu.ec/handle/123456789/2371>
- Vargas Salazar, G. (2017). *El desarrollo del pensamiento lógico-matemático en los estudiantes costarricenses de undécimo año de colegios académicos diurnos y su nivel de logro en el aprendizaje de las matemáticas*. <https://repositorio.uned.ac.cr/xmlui/handle/120809/1646>
- Yasig Salguero, A. de los Á. (2021). *Estrategias metodológicas para desarrollar el razonamiento lógico matemático de los estudiantes de octavo año de la Unidad Educativa César Augusto Salazar Chávez*.

Generación de Texto Guía para la Detección Automatizada del Acoso y el Ciberacoso

Text Generation Guide for Automated Detection of Bullying and Cyberbullying

Marcos Orellana¹ <https://orcid.org/0000-0002-3671-9362>, Jorge Luis Zambrano-Martinez¹ <https://orcid.org/0000-0002-5339-7860>, Ronny Marcelo Calle Andrade¹ <https://orcid.org/0009-0004-9825-155X>,
Amanda Roldan¹ <https://orcid.org/0009-0009-3548-2269>, Andrés Nicolas Tirado Jarama¹ <https://orcid.org/0009-0002-2269-4608>

¹Laboratorio de Investigación y Desarrollo en Informática (LIDI)
Universidad del Azuay, Cuenca, Ecuador

marore@uazuay.edu.ec, jorge.zambrano@uazuay.edu.ec,
ronnycalle@es.uazuay.edu.ec, amy.3107@es.uazuay.edu.ec,
andresnicolas1231@es.uazuay.edu.ec



Esta obra está bajo una licencia internacional
Creative Commons Atribución-NoComercial 4.0

Enviado: 2023/07/14

Aceptado: 2023/08/22

Publicado: 2023/10/15

Resumen

En la actualidad, el acoso entre los niños es un grave problema social e incluso actúa como uno de los principales problemas de la salud mental. Por lo tanto, se han detectado varios tipos de acoso, entre ellos están el acoso físico, verbal, social y el cibernético que utiliza la tecnología al alcance del agresor para difundir el daño y hacer que la víctima esté constantemente acosada. La inteligencia artificial y el acoso son temas ampliamente estudiados e investigados, pero regularmente no son utilizados de manera conjunta para llegar a mitigar la problemática del acoso. Con esto en mente, el propósito de esta investigación es identificar las palabras claves que son extraídas de diferentes medios audiovisuales cuyo tema principal es el acoso y el ciberacoso para crear campos semánticos correspondientes al texto extraído del video y generar un texto guía para la detección automatizada del acoso y el ciberacoso como apoyo a los profesionales.

Palabras clave: Acoso, Ciberacoso, Inteligencia artificial, OpenAI, YAKE.

Sumario: Introducción, Trabajos Relacionados, Materiales y Métodos, Resultados y Conclusiones.

Como citar: Orellana, M., Zambrano-Martinez, J. L., Calle, R. M., Roldan, A. & Tirado, A. N. (2023). Generación de Texto Guía para la Detección Automatizada del Acoso y el Ciberacoso. *Revista Tecnológica - Espol*, 35(2), 181-191. Recuperado a partir de <http://www.rte.espol.edu.ec/index.php/tecnologica/article/view/1049>

Abstract

Currently, bullying among children is a serious social problem and even acts as one of the primary mental health problems. Therefore, various types of bullying have been detected, including physical, verbal, social, and cyberbullying, which uses technology within reach of the aggressor to broadcast the damage and make the victim constantly harassed. Artificial intelligence and bullying are widely studied and researched, but they are not regularly used together to mitigate the bullying problem. Nevertheless, this research aims to identify the keywords extracted from different audiovisual media whose central theme is bullying and cyberbullying to create semantic fields corresponding to the text extracted from the video and generate a guide for automated bullying detection as support for professionals.

Keywords: Bullying, Cyberbullying, Artificial intelligence, OpenAI, YAKE.

Introducción

El acoso o comúnmente conocido en la actualidad como “bullying” se ha convertido en uno de los temas que más preocupa a los padres de familia, educadores de las instituciones y psicólogos, debido a que se ha incrementado el acoso a las víctimas sin distinción de edad, género o nacionalidad, afectando tanto en el ámbito social como económico.

Las víctimas de acoso tienen mayor probabilidad de presentar dificultades con la interacción social en el entorno y ser internadas en centros de salud mental por experimentar problemas de ansiedad, depresión, ideación e intentos suicidas (Heikkilä et al., 2013; Lucas-Molina et al., 2022). Los jóvenes que acosan a otras personas demuestran conductas más inapropiadas que aquellos que son intimidados. Por lo tanto, varias investigaciones han estudiado el impacto de la intimidación que suelen presenciar los individuos o espectadores del acoso. Estos espectadores reportan un aumento de la ansiedad, la hostilidad, la depresión, la paranoia, y el aislamiento emocional (Zang et al., 2020; Alrajeh et al., 2021).

Es importante considerar que existen programas enfocados específicamente sobre la intimidación, en algunos países existen leyes para afrontar el acoso que ocurre en distintos sectores, en especial en las escuelas. Estas políticas guían las acciones e intervenciones dentro de la organización estudiantil referentes a la intimidación. Así, en un estudio realizado por Landstedt y Pearson (2014), en el que se invitó a participar a las escuelas secundarias con estudiantes de edades comprendidas entre 12 y 16 años, a realizar una encuesta con el objetivo de determinar el número de víctimas y victimarios. Los resultados determinaron que el 15.1% de los adolescentes fueron víctimas de acoso, el 17.2% informó haber recibido ciberacoso y el 19.6% de los encuestados sufrió tanto acoso como ciberacoso. Otro resultado importante es el acoso en la población femenina, ya que el 59.4% de las mujeres han recibido acoso por la imagen corporal, frente al 26.2% en los hombres.

En la actualidad, la fusión de investigaciones en varios campos, como la informática y las ciencias sociales, ayuda a analizar distintos problemas en la sociedad. Por ello, varios investigadores han comenzado a estudiar los distintos roles de intimidación en redes sociales, internet y correos electrónicos a través del uso de algoritmos de aprendizaje automático tal como lo mencionan (Akhter et al., 2018; Chatzakou et al., 2017; Chelms et al., 2017; Dhungana Sainju et al., 2021; Nikiforos et al., 2020).

Con la ayuda de la inteligencia artificial es posible encontrar patrones de este tipo de comportamiento para llegar a mitigar y reducir el acoso de los atacantes a las víctimas. Por lo que en esta investigación se propone generar textos guías, a través de herramientas que utilizan

técnicas de minería de texto aplicadas a transcripciones de videos-audios de calidad para la creación de material audiovisual que facilite la reducción del acoso y ciberacoso en la sociedad. Para lograr esta propuesta, es necesario desarrollar las siguientes etapas: i) Extracción del texto de los videos alojados en la plataforma YouTube que estén relacionados con el ciberacoso mediante interfaces de programación de aplicaciones (API) especializadas; ii) Extracción de palabras o frases que sean mayormente representativas de los textos extraídos; iii) Creación de campos semánticos para alimentar la API de OpenIA; y iv) Generación del guion con base a la combinación de las palabras o frases.

El artículo se estructura de la siguiente manera: los trabajos relacionados se presentan en la siguiente sección, luego es presentada la metodología utilizada en el trabajo, después se exponen los resultados y en la última sección se presentan las conclusiones y los trabajos futuros.

Trabajos Relacionados

Existen trabajos enfocados al estudio de las características del acoso y ciberacoso con la finalidad de generar modelos o herramientas para combatirlos. Los principales avances se refieren a la detección, clasificación y predicción del ciberacoso en plataformas en línea, como las redes sociales. A continuación, se exponen los estudios relacionados a la generación de herramientas y modelos utilizando varias fuentes de datos.

Dinakar et al. (2011) proponen una novedosa técnica para la detección del ciberacoso, que consiste en descomponer los textos en varios temas sensibles y generar subproblemas de clasificación. Construyeron un corpus por medio de la extracción de 4,500 comentarios de videos controversiales de YouTube para generar modelos de clasificación binarios y multiclases. Luego de comparar los modelos, los de clasificación binaria tuvieron una mayor exactitud que los multiclases, logrando una exactitud del 80%.

Por otra parte, Zhao et al. (2016) desarrollaron un marco de aprendizaje para la detección de ciberacoso en redes sociales, específicamente en Twitter. Para ello, definieron una lista de palabras ofensivas y la ampliaron a través de incrustaciones de palabras (Word Embeddings). Estos datos alimentaron un modelo de clasificación basado en una máquina de vector de soporte (SVM), logrando una precisión de 76.8%. Por otro lado, Singh et al. (2016) implementaron un marco de fusión de información probabilística que utilizó puntajes de confianza y factores referentes al ciberacoso. Utilizaron un subconjunto de datos etiquetado de 4,865 mensajes de Twitter sobre la tendencia a sufrir ciberacoso. Finalmente, generaron tres modelos, un clasificador basado en la mayoría o categoría principal denominado ZeroR, el método de fusión y su modelo de marco de fusión probabilística, este último demostró un 89% de exactitud.

Mahlangu et al. (2018) revisaron distintos métodos en cuanto a la detección automática de ciberacoso, y además, se identificaron algunos factores, categorías y fuentes de datos en internet. La mayor limitación que encontraron los investigadores corresponde a las escasas fuentes de datos provenientes de imágenes o videos, que pueden ser de utilidad para nuevos estudios en la detección de ciberacoso.

Uludasdemir y Kucuk (2019) realizan un estudio sobre la experiencia de los adolescentes con el ciberacoso y el nivel de conciencia de los padres sobre este problema. Por lo que identificaron que el uso del internet por parte de los adolescentes y el uso de las redes sociales presenta un alto índice de ciberacoso y ciber victimización experimentados por los adolescentes en un desconocimiento por parte de los padres sobre este problema.

Rezvani et al. (2020) presentan un método para orquestar los procesos, modelos y resultados en la detección inteligente del ciberacoso. El uso del método para orquestar les permitió maximizar el uso del contexto y aprovechar fuentes externas. Además, utilizaron un modelo de red de memoria a corto-largo plazo (LSTM) que fue alimentado mediante conjuntos de datos de Twitter e Instagram. Como resultado, el modelo presentado por los autores presenta una exactitud del 87%. De igual manera, Gada et al. (2021) presentan una arquitectura basada en redes neuronales convolucionales con redes de memoria a corto-largo plazo (LSTM-CNN) para la detección de ciberacoso. El modelo fue creado a partir de la extracción de textos de Twitter, además generaron un *bot* para la aplicación de Telegram que alerta al usuario al detectar la presencia de palabras ofensivas o groseras. El rendimiento del modelo fue evaluado a través del análisis de la curva característica operativa del receptor (ROC), dando una exactitud de 97%.

Otro enfoque para la detección de ciberacoso es descrito por Caicedo (2023), que desarrolla un método semisupervisado para detectar, clasificar y anotar textos extraídos de entornos digitales mediante un proceso de evaluación automática de su calidad para detectar y brindar apoyo especializado a usuarios de redes sociales que publican textos con contenido suicida.

Los trabajos mencionados describen algunas de las técnicas y modelos utilizados para la detección, predicción y ayuda para el ciberacoso a los profesionales en solucionar este problema social. Además, mencionaron las limitaciones existentes al afrontar estos problemas, por lo que no consiguen la información generada de otras fuentes multimedia, como los videos e imágenes. Sin embargo, no existen estudios que utilicen generadores de textos modernos en la creación de textos guía para videos que promuevan el combate contra el acoso y ciberacoso.

Materiales y Métodos

En esta investigación, la metodología para realizar la generación de guiones base obtenidos a través de la reproducción de videos con temáticas de acoso y ciberacoso, con la finalidad de prevenir y disminuir esta problemática a nivel escolar fue empleada utilizando las siguientes fases: i) Extracción del texto de diferentes videos relacionados con el acoso y ciberacoso, ii) extracción de las palabras más relevantes e importantes a través de la herramienta *Yet Another Keyword Extractor* (YAKE), iii) Construcción de los campos semánticos utilizando el generador de texto de OpenAI y iv) Generación del guion adecuado combinando palabras o frases para el acoso y ciberacoso.

Extracción del texto proveniente de videos seleccionados

Como fase inicial, se tuvo que realizar una búsqueda de los videos en la plataforma YouTube, seleccionando los primeros videos que presentó el algoritmo que utiliza la plataforma y alineados con la temática de este estudio. Luego, se elaboró una videoteca que contenía distintos elementos audiovisuales sobre el acoso y el ciberacoso conjuntamente con la duración y el código con el que se puede encontrar dichos elementos audiovisuales, como se puede apreciar en la Tabla 1.

Para la extracción del texto desde los videos seleccionados se utilizaron tres APIs, YouTube translate, Amberscript y Chequeabot. Estas aplicaciones extraen los textos del elemento audiovisual utilizando algoritmos e inteligencia artificial con una alta precisión para la creación de párrafos con sentido semántico, cuya finalidad es reducir el número de errores posibles al realizar la transcripción del video al texto en varios lenguajes y sus distintas pronunciaciones particulares de individuos de un mismo idioma.

Tabla 1
Videoteca sobre acoso y ciberacoso

NÚMERO VIDEO	TÍTULO	DURACIÓN	CÓDIGO
1	“Cómo evitar el acoso escolar”.	3 min 27 sec.	nTRZIMfprBI
2	“Cómo actuar frente al acoso escolar. José Antonio Luengo, psicólogo y profesor”.	4 min 51 sec.	2Un_J3EhUqk
3	“Cómo prevenir el bullying en el aula”.	3 min 53 sec.	FHzAnNQBzA8
4	“Cyberbullying: qué es y cómo prevenirlo”.	1 min 59 sec.	zdJIsMbwU08
5	“Cyberbullying - ¿Qué es y qué hacer?”.	4 min 22 sec.	Gfsuj6wlg8Y
6	“¿Qué es el Bullying?”.	2 min 07 sec.	YGMpU35TwNY
7	“Cyberbullying, Ciberacoso y depresión”.	1 min 18 sec.	3_e_M0ntZpM
8	“Qué hacer contra el Bullying”.	3 min 06 sec.	AwEqxDtJu0

Sin embargo, los párrafos fueron minuciosamente revisados y supervisados al momento de la extracción por los expertos en temas de acoso y ciberacoso con la finalidad de disminuir las probabilidades de errores semánticos y errores de transcripción.

Extracción de palabras clave

Una vez obtenidos los textos transcritos de los videos, es necesaria la extracción de las palabras claves provenientes de los párrafos creados por los transcriptores. Por este motivo, se utilizó un método de extracción automática de palabras importantes basado en características estadísticas de texto extraídas de textos individuales sin ninguna supervisión para seleccionar las palabras claves más relevantes de un texto denominado YAKE (Campos et al., 2020). Además, esta herramienta no necesita de un entrenamiento previo con un conjunto particular de documentos, ni de diccionarios, o corpus externos (Campos et al., 2018). Con la herramienta YAKE se facilita la extracción de palabras clave que presentan un vínculo con el acoso y el ciberacoso, siendo las palabras de mayor relevancia y frecuencia. Para brindar una mejor interpretación de las palabras claves obtenidas de los videos, se desarrolló un resumen representado en la Tabla 2.

Además, como se observa en la Tabla 2, se obtuvo un puntaje de las palabras clave a través de la Ecuación 1 presentada por Campos et al. (2020):

$$S(kw) = \frac{\prod_{w \in kw} S(w)}{TF(kw) \times (1 + \sum_{w \in kw} S(w))} \quad (1)$$

Donde $S(w)$ es el peso o la puntuación de la palabra y $TF(kw)$ es la frecuencia de la palabra clave que se presenta en el texto analizado acorde a ecuaciones previas del trabajo de Campos et al. (2018). Esta ecuación nos indica cuáles palabras tienen un mayor significado semántico dentro del texto analizado acerca del tema de acoso y ciberacoso, para facilitar la creación de los campos semánticos.

Tabla 2
Matriz de las palabras usadas con mayor frecuencia y relevancia

NÚMERO VIDEO	POSICIÓN	PALABRAS	PUNTAJE	RELEVANTE
1	1	acoso escolar	0,02346838	Si
	2	escuela	0,051890499	No
	2	acoso	0,058818566	Si
	4	sociedad	0,07814722	No
	5	maltrato	0,081237283	Si
...
8	1	tus amigos recuerdan	0,000132623	No
	2	agresiones físicas seguramente	0,000202302	Si
	3	obtener ayuda profesional	0,000202302	No
	4	apodosos insultos amenazas	0,000242478	Si
	5	hacerte sentir insignificante	0,000242478	Si

Como YAKE extrae las palabras claves basado en estadísticas, existen palabras repetidas, por lo que fue necesario aplicar una función de *deduplicación de datos* para la eliminación de duplicados exclusivos de los datos y reducirlos. Entre ellos se tiene *distancia de Levenshtein*, *distancia de Jaro-Winkler* y *Sequence Matcher*. Además, otros parámetros para la extracción de las palabras importantes fueron el máximo tamaño de n-gramas, el lenguaje del texto, el límite de la puntuación normalizada de la función de *deduplicación*, y el número máximo de palabras clave.

Construcción de los cuerpos semánticos

Esta fase corresponde a la creación de los campos semánticos una vez que se ha completado la extracción de las palabras clave. Para ello, se seleccionaron las palabras clave más relevantes a través de la evaluación de sus puntajes, tomando en cuenta que, a menor puntuación, mayor es la relevancia de la palabra dentro del texto. Adicionalmente, algunas palabras fueron filtradas por un experto en temas de acoso y ciberacoso, debido a que presentaban una baja relevancia en la construcción de los cuerpos semánticos.

Una vez que las palabras claves fueron seleccionadas, se utilizó el módulo generativo de texto de la compañía OpenAI. Esta empresa se caracteriza por promover y desarrollar productos relacionados con Inteligencia Artificial (IA) como transformadores pre-entrenados generativos (GPT).

Este módulo de OpenAI es capaz de generar textos a través de palabras o porciones de texto proporcionadas por el usuario. Además, existen varios modelos disponibles a diferentes precios y capacidades, siendo así *Text-Davinci-003*, el más sofisticado, preciso, costoso y basado en la familia GPT versión 3.5. Este modelo se especializa en tareas de lenguaje natural como la traducción y la generación textual. Además, se modificaron algunos de los parámetros más relevantes dentro del módulo de generación de textos, como la temperatura, el número máximo de tokens y el modo, la finalidad de estos parámetros se describe en la Tabla 3. También existen otros modelos con menor precisión, pero más rápidos y económicos, como *Text-Curie-001* y *Text-Babbage-001* (Efrat et al. 2022). Además de seleccionar el modelo Text-Davinci-003.

Tabla 3
Parámetros del módulo generativo de textos

PARAMETROS	DESCRIPCIÓN
Temperatura	Oscila entre 0 y 1, a medida que aumenta el valor, aumenta la creatividad de la respuesta.
Número de tokens	El máximo número de tokens a generar, entre 1 y 4000 tokens permitidos para el modelo Text-Davinci-003.
Modo	Especifica la tarea que va a realizar el modelo: completar, responder, agregar o editar texto.

En la Figura 1 se evidencian los cuerpos semánticos generados por el módulo de OpenAI que puede ser implementado mediante un lenguaje de programación a través de una API o a través de su plataforma en línea. Los cuerpos semánticos generados fueron evaluados por expertos en las temáticas de acoso y ciberacoso para verificar su contenido y consistencia.

Figura 1
Construcción de campos semánticos en el módulo de OpenAI



Generación del guion combinando frases o palabras extraídas

Una vez que las palabras claves y los cuerpos semánticos fueron validados por el experto, como fase final de la metodología realizada para este estudio, se generó el guion para el combate del acoso y ciberacoso a través de los cuerpos semánticos generados en la fase anterior. Para la generación del guion se desarrolló un script en el lenguaje de programación Python que concatena las palabras claves con sus correspondientes cuerpos semánticos a través de las APIs descritas anteriormente.

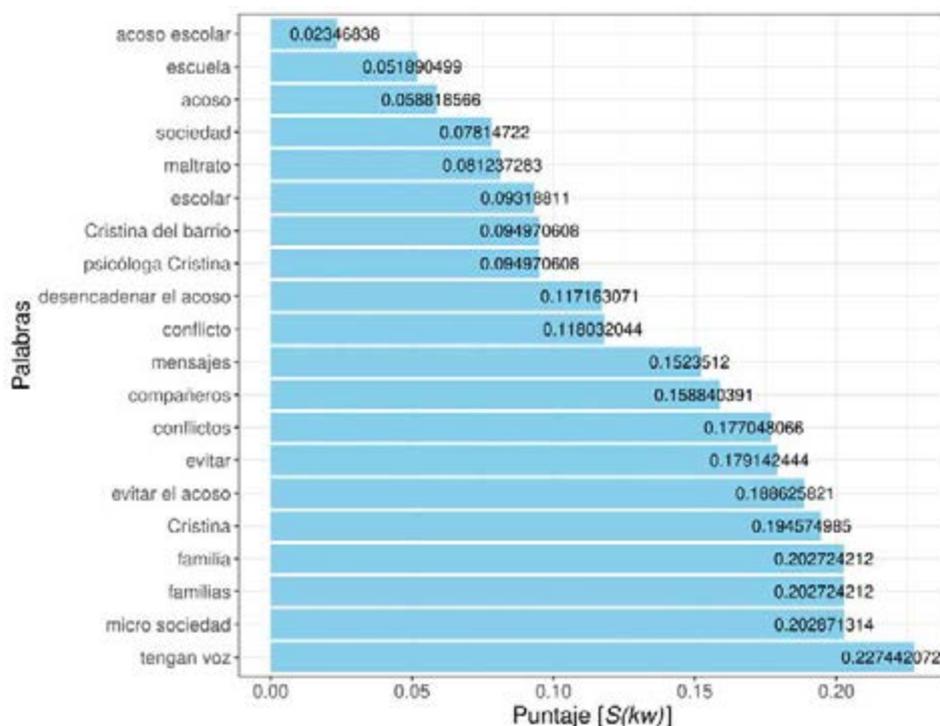
Resultados

En los resultados se encontró que, la relación entre el acoso o el ciberacoso, con la extracción de datos o variables para su posterior procesamiento aplicando IA son muy escasas. Sin embargo, se logra rescatar algunos documentos donde se evidencia esta relación, y esto brindó los indicios para la elaboración de esta investigación. Esto se debe a que la vida humana se está trasladando cada vez a lo virtual, presentando menos frecuencia de un acoso físico y la facilidad del mundo tecnológico para realizar acosos.

El resultado primordial en este estudio es la creación de un texto guía para la elaboración de un guion de video contra el acoso y el ciberacoso. Para lo cual, primero se realizó la selección de algunos videos que explican sobre el tema de acoso y ciberacoso, y luego fue transcrito el texto completo de los videos seleccionados con herramientas especializadas. Después, se extrajeron las palabras claves con un sistema basado en características para la extracción de las palabras claves plurilingües y seguidamente, la construcción de los cuerpos semánticos a través de las palabras o frases con mayor relevancia del texto.

Así pues, el algoritmo YAKE que extrae las palabras claves, requirió de parámetros como el idioma en español, el tamaño de las gramas de tres a cinco, con un algoritmo de *deduplicación Sequence Meter* de un límite igual a 0.9, y que presente las primeras 20 palabras obtenidas. Las palabras claves candidatas fueron mostradas con un peso o puntaje cuya mayor relevancia para el texto procesado es la menor posible, tal como se puede observar en la Figura 2.

Figura 2
Palabras claves extraídas de un texto transcrito de video



Asimismo, se tuvo la necesidad de realizar un filtrado por parte de expertos en el tema a las palabras claves que fueron extraídas de las transcripciones. Esto es debido a palabras o frases que no presentan una relevancia para la generación del texto guía como “sociedad”, “escolar”, “Cristina del barrio”, entre otros, que, aunque presenten un puntaje inferior marcado, no destacan para el estudio de esta investigación. Por ende, se redujo la cantidad de palabras importantes extraídas para la creación de los cuerpos semánticos y su posterior generación del texto guía.

No obstante, para la generación del guion para el combate del acoso y el ciberacoso se desarrolló un script en Python que permitió la agrupación correcta de las palabras clave con sus propios cuerpos semánticos. Sin embargo, al presentarse una gran cantidad de datos relacionados al acoso y ciberacoso en la creación de texto guía, se tuvo que refinar por los expertos para encaminar el contexto del guion y que esté centrado directamente en la prevención del acoso y ciberacoso. En la Figura 3, se puede observar el guion generado utilizando las palabras claves y los cuerpos semánticos obtenidos con anterioridad.

Figura 3

Generación del texto guía a través de los campos semánticos con OpenAI

```

t141@raspberrypi:~/Desktop$ python3 generacion_texto_guia.py

Definición Generada:

User: Cyberbullying es violencia
System: El ciberacoso o cyberbullying es una forma de violencia en línea que afecta tanto a niños como a adultos. Este tipo de agresión puede tener consecuencias devastadoras para sus víctimas, incluyendo ansiedad, depresión, aislamiento social e incluso aumento del riesgo de suicidio. La violencia en línea incluye todo tipo de comportamiento abusivo, como insultos, difamación, exposición de información privada y amenazas. Es importante tomar medidas para combatir el ciberacoso y crear una cultura en línea más segura y respetuosa. Los niños y adultos deben aprender a manejar la tecnología para que no se convierta en una herramienta para el acoso y la violencia.

User: Consecuencias graves de violencia
System: La violencia tiene consecuencias graves tanto para las víctimas como para los agresores y la sociedad en general. Las víctimas pueden sufrir lesiones físicas y emocionales, trastornos psicológicos, dificultades para relacionarse, pérdida de confianza y autoestima, y en los casos más extremos, la muerte. Los agresores pueden enfrentarse a cargos criminales, problemas legales, sentimientos de culpa y vergüenza, y dificultades para reintegrarse a la sociedad. La violencia también tiene un impacto negativo en la comunidad, generando miedo, desconfianza, división y dañando la convivencia pacífica.

User: Impedir que siga ocurriendo
System: Para evitar que ocurran más, es importante identificar la causa raíz del problema e implementar medidas para abordarlo. Esto puede implicar cambios en las políticas, los procedimientos o la capacitación de las personas involucradas. El monitoreo y la revisión regulares de la situación también pueden ayudar a identificar cualquier problema nuevo que pueda surgir y tomar medidas correctivas rápidamente. Es crucial establecer una cultura de rendición de cuentas y transparencia para garantizar que todos sean conscientes de sus responsabilidades y las consecuencias de las acciones inapropiadas. La educación y la comunicación continuas pueden ayudar a reforzar estos valores y evitar que ocurran incidentes en el futuro.

User: Víctima de cyberbullying
System: Ser víctima de ciberacoso puede ser una experiencia aterradora y traumática. Los insultos, rumores y amenazas que se pueden recibir en línea pueden tener un impacto negativo en la autoestima, la salud mental y el rendimiento en la escuela o el trabajo. Es importante hablar sobre ello con alguien en quien se confía, como un amigo, un familiar o un profesional de la salud mental. También es útil aprender a gestionar la privacidad en línea y a filtrar los mensajes negativos para evitar que afecten a uno de manera personal.

User: Evitar violencia
System: Para evitar la violencia, es importante fomentar la tolerancia, el diálogo y la empatía. Es fundamental educar desde temprana edad en valores de respeto, paz y convivencia pacífica. Asimismo, es necesario combatir las desigualdades sociales, económicas y culturales, ya que estas suelen ser causantes de la violencia. En lugar de recurrir a la violencia como solución a los conflictos, es importante buscar alternativas pacíficas como la negociación y la mediación. También es importante la participación ciudadana en la prevención y la denuncia de la violencia.

User: Consecuencias que los agresores sufren
System: Los agresores pueden experimentar diferentes tipos de consecuencias a raíz de su comportamiento violento, como problemas legales y penales, exclusión social y aislamiento, sentimientos de culpa y vergüenza, y deterioro de las relaciones personales. También pueden presentar problemas de salud mental, como ansiedad, depresión y trastornos de personalidad. En algunos casos, los agresores pueden darse cuenta del daño que han causado y buscar ayuda para abordar sus comportamientos violentos y trabajar en su rehabilitación.

User: Como mejorar si es un agresor
System: Si una persona se identifica como un agresor, lo primero que debe hacer es buscar ayuda profesional. Es importante hablar con un terapeuta o psicólogo para comprender la raíz de su comportamiento agresivo y aprender técnicas adecuadas para controlar sus emociones y resolver conflictos de manera pacífica. También es importante asumir la responsabilidad por su comportamiento y trabajar para reparar cualquier daño causado a las víctimas. Finalmente, debe comprometerse a cambiar su comportamiento y actuar de manera respetuosa y no violenta en el futuro.

User: 

```

Conclusiones

Para concluir, los avances en materia de inteligencia artificial han demostrado su capacidad para afrontar problemas y retos de la sociedad actual, brindando soluciones efectivas en diversos temas como el acoso y el ciberacoso. De acuerdo a estudios previos, son escasas las investigaciones que utilizan fuentes de datos provenientes de materiales audiovisuales; es así que en este trabajo se resaltó el uso de datos obtenidos a través de transcripciones de videos relacionados al acoso y ciberacoso. El guion generado en esta investigación pretende ser un apoyo para la creación de futuros guiones para combatir el acoso y el ciberacoso.

De este modo, se utilizó el módulo de generación de texto de la compañía OpenAI en la construcción de un guion de video para combatir el acoso y el ciberacoso, utilizando datos de videos relacionados a estas temáticas. Además, el uso de la herramienta YAKE simplificó los procesos más tradicionales de minería de textos, facilitando la extracción de palabras claves. No obstante, fue necesaria la intervención de expertos en el tema de acoso y ciberacoso para comprobar tanto las palabras claves como el guion generado, y evitar inconsistencias semánticas en el guion generado.

Como trabajos futuros, se pretende extraer más información de videos relacionados con el acoso y ciberacoso, generar más cuerpos semánticos y construir un corpus que pueda utilizarse en distintos modelos de aprendizaje automático y aprendizaje profundo. Y, finalmente, que dichos modelos puedan ser implementados en centros educativos para brindar apoyo a los profesionales, donde la presencia de acoso y ciberacoso es mayor.

Reconocimientos

Los autores desean agradecer al Vicerrectorado de Investigaciones de la Universidad del Azuay por el apoyo financiero y académico, así como a todo el personal de la escuela de Ingeniería de Ciencias de la Computación, y el Laboratorio de Investigación y Desarrollo en Informática (LIDI).

Referencias

- Alrajeh, S., Hassan, H., Al-Ahmed, A., y Alsayed Hassan, D. (2021). An investigation of the relationship between cyberbullying, cybervictimization and depression symptoms: A cross sectional study among university students in Qatar. *PLoS one*, 16(12), e0260263.
- Akhter, S., y others (2018). Social media bullying detection using machine learning on Bangla text. In *2018 10th International Conference on Electrical and Computer Engineering (ICECE)* (pp. 385–388).
- Caicedo, R. (2023). Metodo Semi-Supervisado para Detectar, Clasificar y Anotar en un Corpus de Suicidio Textos extraídos de Entornos Digitales. (*Doctoral dissertation, Pontificia Universidad Católica del Perú (Peru)*).
- Campos, R., Mangaravite, V., Pasquali, A., Jorge, A. M., Nunes, C., y Jatowt, A. (2018). YAKE! Collection-Independent Automatic Keyword Extractor (pp. 806-810). https://doi.org/10.1007/978-3-319-76941-7_80
- Campos, R., Mangaravite, V., Pasquali, A., Jorge, A., Nunes, C., y Jatowt, A. (2020). YAKE! Keyword extraction from single documents using multiple local features. *Information Sciences*, 509, 257-289. <https://doi.org/10.1016/j.ins.2019.09.013>
- Chatzakou, D., Kourtellis, N., Blackburn, J., De Cristofaro, E., Stringhini, G., y Vakali, A. (2017). Mean Birds. *Proceedings of the 2017 ACM on Web Science Conference*, 13-22. <https://doi.org/10.1145/3091478.3091487>
- Chelmiss, C., Zois, D.-S., y Yao, M. (2017). Mining Patterns of Cyberbullying on Twitter. *2017 IEEE International Conference on Data Mining Workshops (ICDMW)*, 126-133. <https://doi.org/10.1109/ICDMW.2017.22>
- Dhungana Sainju, K., Mishra, N., Kuffour, A., y Young, L. (2021). Bullying discourse on Twitter: An examination of bully-related tweets using supervised machine learning. *Computers in Human Behavior*, 120, 106735. <https://doi.org/10.1016/j.chb.2021.106735>
- Dinakar, K., Reichart, R., y Lieberman, H. (2011). Modeling the detection of textual cyberbullying. *Proceedings of the International AAAI Conference on Web and Social Media*, 5(3), 11-17.
- Efrat, A., Honovich, O., & Levy, O. (2022). Lmentry: A language model benchmark of elementary language tasks. *arXiv preprint arXiv:2211.02069*.
- Gada, M., Damania, K., y Sankhe, S. (2021). Cyberbullying detection using lstm-cnn architecture and its applications. *2021 International Conference on Computer Communication and Informatics (ICCCI)*, 1-6.
- Heikkilä, H.-K., Väänänen, J., Helminen, M., Fröjd, S., Marttunen, M., y Kaltiala-Heino, R. (2013). Involvement in bullying and suicidal ideation in middle adolescence: a 2-year follow-up study. *European Child y Adolescent Psychiatry*, 22(2), 95-102. <https://doi.org/10.1007/s00787-012-0327-0>
- Landstedt, E., y Persson, S. (2014). Bullying, cyberbullying, and mental health in young people. *Scandinavian Journal of Public Health*, 42(4), 393-399. <https://doi.org/10.1177/1403494814525004>
- Lucas-Molina, B., Pérez-Albéniz, A., Solbes-Canales, I., Ortuño-Sierra, J., y Fonseca-Pedrero, E. (2022). Acoso escolar, cyberbullying y salud mental: el rol de la cohesión entre estudiantes como factor protector escolar. *Psychosocial Intervention*, 31(1), 33–41s
- Mahlangu, T., Tu, C., y Owolawi, P. (2018). A review of automated detection methods for cyberbullying. *2018 International Conference on Intelligent and Innovative Computing Applications (ICONIC)*, 1-5.
- Nikiforos, S., Tzanavaris, S., y Kermanidis, K.L. (2020). Virtual learning communities (VLCs) rethinking: influence on behavior modification—bullying detection through machine learning and natural language processing. *Journal of Computers in Education*, 7, 531–551.
- Rezvani, N., Beheshti, A., y Tabebordbar, A. (2020). Linking textual and contextual features for intelligent cyberbullying detection in social media. *Proceedings of the 18th International Conference on Advances in Mobile Computing y Multimedia*, 3-10.

- Rivers, I., Poteat, V. P., Noret, N., y Ashurst, N. (2009). Observing bullying at school: The mental health implications of witness status. *School Psychology Quarterly*, 24(4), 211-223. <https://doi.org/10.1037/a0018164>
- Singh, V. K., Huang, Q., y Atrey, P. K. (2016). Cyberbullying detection using probabilistic socio-textual information fusion. *2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, 884-887.
- Uludasdemir, D., y Kucuk, S. (2019). Cyber bullying experiences of adolescents and parental awareness: Turkish example. *Journal of pediatric nursing*, 44, e84–e90.
- Zhang, D., Huebner, E., y Tian, L. (2020). Longitudinal associations among neuroticism, depression, and cyberbullying in early adolescents. *Computers in Human Behavior*, 112, 106475.
- Zhao, R., Zhou, A., y Mao, K. (2016). Automatic detection of cyberbullying on social networks based on bullying features. *Proceedings of the 17th international conference on distributed computing and networking*, 1-6.

Síntesis Textual de Evaluación para Acoso y Ciberacoso

Textual Synthesis of Test for Bullying and Cyberbullying

Marcos Orellana¹ <https://orcid.org/0000-0002-3671-9362>, Jorge Luis Zambrano-Martinez¹ <https://orcid.org/0000-0002-5339-7860>, Patricio Santiago Garcia Montero¹ <https://orcid.org/0009-0007-4113-8400>,
Liliana Marilu Lojano¹ <https://orcid.org/0009-0006-4993-8747>, Mateo Sebastian Zea¹ <https://orcid.org/0009-0005-4209-8143>, Tupak Pacjakutik Japon¹ <https://orcid.org/0009-0005-3239-992X>

¹Laboratorio de Investigación y Desarrollo en Informática (LIDI)
Universidad del Azuay, Cuenca, Ecuador

marore@uazuay.edu.ec, jorge.zambrano@uazuay.edu.ec,
santyg20@es.uazuay.edu.ec, liliana1998@es.uazuay.edu.ec,
mzea582@es.uazuay.edu.ec, tupak.japon@es.uazuay.edu.ec



Esta obra está bajo una licencia internacional
Creative Commons Atribución-NoComercial 4.0

Enviado: 2023/07/14

Aceptado: 2023/08/15

Publicado: 2023/10/15

Resumen

En los últimos años, el acoso y el ciberacoso son problemas que han aumentado vertiginosamente afectando escuelas, colegios y universidades. Debido a los avances en las tecnologías de la información, cualquier persona está expuesta a ser atacada; por esta razón, es necesario crear soluciones a través de técnicas adecuadas que ayuden a prevenir el acoso y ciberacoso. En consecuencia, en este artículo se propone crear una síntesis textual a partir de datos de encuestas que permita desarrollar modelos para clasificar o predecir tanto a víctimas como agresores de acoso y ciberacoso. Para ello, se utilizaron técnicas de minería de datos, árboles de decisión y técnicas de agrupación, dando como resultado una síntesis textual. Esto permitió la creación y evaluación de un modelo de aprendizaje supervisado y otro modelo con técnicas de agrupamiento, aplicadas a los datos de las encuestas realizadas a estudiantes universitarios. Los resultados demostraron la importancia de la síntesis textual para la generación de modelos de clasificación o predicción de víctimas y agresores del acoso y ciberacoso, con una exactitud mayor al 75%, siendo el modelo de agrupamiento con mejor rendimiento.

Palabras clave: Acoso, Ciberacoso, Minería de datos, Árbol de decisión, Agrupamiento.

Sumario: Introducción, Trabajos relacionados, Marco Teórico, Materiales y Métodos, Resultados y Discusión y Conclusiones.

Como citar: Orellana, M., Zambrano-Martinez, J. L., Garcia, P. S., Lojano, L. M., Zea, M. S. & Japon, T. P. (2023). Síntesis Textual de Evaluación para Acoso y Ciberacoso. *Revista Tecnológica - Espol*, 35(2), 192-205. Recuperado a partir de <http://www.rte.espol.edu.ec/index.php/tecnologica/article/view/1050>

Abstract

In recent years, bullying and cyberbullying have increased, affecting schools, colleges, and universities. Due to advances in information technology, any person is exposed to being attacked. Therefore, it is necessary to create solutions through appropriate techniques that help prevent bullying and cyberbullying. Consequently, this article proposes to create a textual synthesis from survey data that allows the development of models to classify or predict both victims and aggressors of bullying and cyberbullying. Data mining techniques, decision trees, and grouping techniques were used, resulting in a textual synthesis. This allowed the creation and evaluation of a supervised learning model and another model with clustering techniques applied to the data from the surveys performed on university students. The results demonstrated the importance of textual synthesis for generating models for the classification or prediction of victims and aggressors of bullying and cyberbullying, with an accuracy greater than 75% for the grouping model with the best performance.

Keywords: Bullying, Cyberbullying, Data mining, Decision tree, Clustering.

Introducción

La violencia es un fenómeno global que afecta a la sociedad en varios escenarios de la vida cotidiana, como los hogares, colegios, lugares de trabajo y relaciones personales. Su presencia trae consecuencias negativas en cualquier ámbito, ya sea social, económico, político o gubernamental. No tiene distinción por edad, género, raza, o cultura y puede expresarse de varias formas. Además, tiene un impacto devastador en la vida de la persona que la sufre, alterando aspectos de su salud física y mental. La violencia puede ser verbal, física, sexual, psicológica o de género (Nielsen y Einarsen, 2018). No obstante, una forma que se encuentra a menudo en la sociedad moderna es el acoso, un tipo de violencia que puede ir desde el acoso verbal hasta la agresión física o psicológica e incluso digital, como el ciberacoso (Mollo et al., 2018).

El auge de las redes sociales, juegos en línea y aplicaciones de mensajería instantánea ha evolucionado la forma en que las personas interactúan, sin embargo, han dado lugar a una creciente forma de violencia. El ciberacoso utiliza cualquier medio de difusión digital para acosar, intimidar, chantajear o difamar a una persona, a diferencia del acoso tradicional, que utiliza medios físicos o verbales para el mismo fin. Ambos tipos de violencia están presentes en los mismos escenarios, pero el ciberacoso posee un mayor alcance debido a las crecientes redes digitales y de telecomunicaciones que conectan en la actualidad a la mayoría de la población (Castilla, 2021). Sin embargo, toda esa interacción digital deja grandes cantidades de información que puede ser explorada y explotada mediante técnicas y algoritmos de minería de datos para combatir y mitigar estas formas de violencia.

La minería de datos permite analizar información con el fin de procesar y explorar cualquier patrón o relación no evidente a simple vista (Han et al., 2022). En este estudio se determinó el uso de la minería de datos como una herramienta para analizar y sintetizar los datos de una encuesta aplicada a estudiantes universitarios sobre sus experiencias con el acoso y ciberacoso, aplicando técnicas como el preprocesamiento de datos, técnicas de agrupamiento y modelos de clasificación. La finalidad del estudio es generar una síntesis textual de las ideas primordiales obtenidas de las encuestas que actúe como punto de partida para la generación de conocimiento sobre el acoso y el ciberacoso.

El resto del documento se estructura de la siguiente manera: Sección 2, presenta los trabajos relacionados, la Sección 3 plantea los fundamentos teóricos de la investigación, la

Sección 4 expone los materiales y métodos utilizados, la Sección 5 explica los resultados que se han obtenido luego de aplicar los métodos y la Sección 6 presenta las conclusiones y trabajos futuros.

Trabajos relacionados

La presencia del acoso y el ciberacoso genera gran discusión en la comunidad académica, lo que ha llevado a varios investigadores a trabajar para beneficio de las personas que padecen a diario estas agresiones (Arce-Ruelas et al., 2022). Los trabajos que se exponen a continuación demuestran el uso de la minería de datos y textos en la generación de modelos de detección y clasificación de estos fenómenos.

En el trabajo de Shaikh et al. (2020), desarrollan una revisión sistemática de literatura para identificar los factores que impulsan a estudiantes universitarios hacia el ciberacoso. Su trabajo pretende servir de guía para futuras investigaciones en el análisis del ciberacoso. Identificaron cerca de 35 factores en 32 estudios, siendo los más reportados problemas emocionales (depresión, ansiedad y estrés), autoestima, agresión, personalidad, malas relaciones, estilo de crianza, rendimiento académico, falta de empatía, exposición tecnológica y facilidad de acceso a internet.

Del mismo modo, Namane y Kyobe (2017), a través de la Universidad de Cape Town, relataron el desarrollo de un análisis con respecto al comportamiento y características de las personas agresoras. En el análisis participaron 3,621 personas (víctimas y agresores) con un promedio de edad entre 14 y 18 años. A través de una encuesta, los investigadores demostraron que 407 participantes fueron víctimas de acoso y se separaron en tres índices de riesgo social. Estos índices se comprenden por la probabilidad que una persona sufra de algún perjuicio dentro del entorno que le rodea (Jorgensen y Siegel, 2019). Así, según Namane y Kyobe (2017), 107 participantes fueron víctimas de acoso provenientes de sectores con bajo riesgo social, 114 participantes se ubicaron dentro del medio riesgo social y 186 participantes dentro del alto riesgo social sufrieron acoso. Además, enfatizaron que los comportamientos violentos se presentan con mayor frecuencia en adolescentes, debido a que es una etapa con mucha variación y cambios constantes en la personalidad, actitudes y emociones.

Por lo tanto, los adolescentes se sitúan como uno de los mayores grupos de consumidores y creadores de contenido en las redes sociales, así en el trabajo de (Bozyiğit et al., 2019) se evaluaron ocho modelos distintos de redes neuronales para la detección de ciberacoso en tweets de Turquía. Los investigadores emplearon técnicas de minería de textos para procesar la información como la tokenización, transformación y la eliminación de símbolos, en conjunto con otras técnicas como frecuencia de término – frecuencia inversa de documento (TF-IDF) y N-gramas para procesar 3,000 tweets y entrenar los distintos modelos. Finalmente, solo un modelo presentó una exactitud del 91%, debido a la utilización de varias configuraciones. Los investigadores también demostraron que incrementar el número de capas ocultas no necesariamente mejora el rendimiento del modelo.

Adicionalmente, para implementar un tipo de red neuronal denominada memoria a corto-largo plazo (LSTM) en el trabajo de Mahat (2021), los investigadores utilizaron datos de Twitter, Wikipedia y Formspring. Un total de 9,000 registros fueron procesados con técnicas de minería de textos (eliminación de caracteres especiales, espacios y ruido en general). Al final el modelo tuvo una exactitud del 77.9%. Otra aplicación de modelos de aprendizaje profundo en redes sociales se aprecia en el trabajo de Banerjee et al. (2019), donde utiliza un modelo basado

en una red neuronal convolucional para la detección del ciberacoso en India. Cerca de 69,874 tweets fueron extraídos y procesados para eliminar palabras vacías y signos de puntuación. Posteriormente, los textos fueron vectorizados para alimentar al modelo, que logró una exactitud de 93.97%, superando a otros modelos de aprendizaje automático, como la máquina de vector de soporte (SVM).

En contraste con lo anterior, Dalvi et al. (2020) presenta un método para la detección de ciberacoso que utiliza una SVM y un clasificador Naive Bayes. Recuperaron tweets de locaciones en tiempo real y utilizaron técnicas de preprocesamiento de datos (tokenización, eliminación de signos de puntuación y palabras vacías, lematización y transformación) a través de un paquete de herramientas de lenguaje natural (NLTK). Al final, el algoritmo SVM es el que presenta una mayor exactitud, con el 71.25%, a diferencia del clasificador bayesiano, con 52.70%.

Si bien las redes sociales son utilizadas a menudo en el análisis del acoso y ciberacoso, también existen estudios enfocados a detectarlos en otras plataformas, como los juegos en línea. Cornel et al. (2019) desarrollan una red neuronal convolucional para detectar la presencia de ciberacoso en los registros de mensajería de dos juegos en línea: *Dota* y *Ragnarok*. A través de varias interfaces de programación de aplicaciones (APIs) recogieron 230,394 frases de *Dota* pertenecientes a usuarios de Filipinas y 534,328 de usuarios que jugaron *Ragnarok* en Japón y Singapur. Luego de eliminar las palabras vacías y caracteres especiales, vectorizaron los registros a través del algoritmo *word embedding* para alimentar al modelo. La red neuronal tuvo una exactitud de 99.93%, sin embargo, los investigadores concluyeron que el modelo tiende a sobre ajustarse, por lo que recomendaron explorar otros modelos de aprendizaje profundo.

Como se evidencia en los trabajos mencionados, las redes neuronales son utilizadas a menudo en el análisis de estos fenómenos utilizando distintos modelos. De igual forma, Rahman et al. (2021) evalúan los siguientes modelos de aprendizaje automático: SVM, árboles de decisión, bosque aleatorio, regresión logística y clasificador bayesiano. A través de varias plataformas (Kaggle, Twitter, Wikipedia y YouTube) recuperaron 31,403 registros etiquetados como inofensivos, y 23,663 etiquetados como ciberacoso. Para este análisis, se utilizan técnicas de preprocesamiento de datos (limpieza, derivación, eliminación de ruido y palabras vacías) y el valor TF-IDF para construir los modelos. Como resultado, el bosque aleatorio presentó la mayor precisión de todos, con el 89%.

En base a los estudios mencionados, los avances en la detección y clasificación del acoso y ciberacoso van desde el análisis de comportamientos hasta la aplicación de modelos de aprendizaje profundo. Todos ellos emplearon técnicas de minería de datos o textos y algunos de ellos coincidieron en varias tareas de preprocesamiento. Sin embargo, la tarea de síntesis previa de los datos es fundamental, debido a que los datos suelen contener ruido y afectan al rendimiento de los modelos (Tapia et al., 2018). Por lo tanto, son escasos los trabajos que resaltan la importancia y el uso de diferentes técnicas en la síntesis de texto en el análisis del acoso y el ciberacoso.

Marco Teórico

El acoso y el ciberacoso son temas extensos por tratar y abarcan varios campos de la psicología, desde conductas humanas hasta rasgos de personalidad que surgen en edades tempranas, como la niñez y adolescencia. Cuando se presenta, suele repetirse a menudo en ambientes donde el agresor siente confianza y que el agredido suele frecuentar como las

escuelas, colegios, universidades y demás instituciones educativas. Las agresiones van desde insultos, golpes, discriminaciones, burlas, hasta publicaciones de información, fotografías y videos íntimos que buscan difamar y avergonzar a la persona agredida. El ciberacoso utiliza la tecnología como su herramienta principal para agredir de manera pública o anónimamente a su víctima desde cualquier parte del mundo. Estas agresiones pueden incurrir en varios problemas físicos y mentales de la persona agredida, desde falta de confianza, problemas para socializar, depresión hasta lesiones físicas o suicidios (Li et al., 2022).

Tapia et al. (2018) describen al acoso como una conducta intencional carente de ética, inmoral e impropia, que se basa en una serie de amenazas físicas o verbales hacia otra persona, generando angustia en la víctima y un desequilibrio de poder entre su persona y el agresor. Mientras tanto, el ciberacoso es una amenaza que afecta a la sociedad moderna gracias al surgimiento de nuevas tecnologías de información. Si bien el avance tecnológico ha sido de beneficio en áreas críticas como la salud y la educación, también ha generado un problema social grave. El anonimato en la era digital en conjunto con problemas o trastornos de una persona puede resultar en posibles agresores, que perciben la violencia como una salida a su conflictiva realidad. En contraste, la persona agredida tiende a dificultar la comunicación con su entorno, lo que reduce la posibilidad de que algún familiar o una persona de su círculo cercano logre identificar la presencia de ciberacoso fácilmente.

De acuerdo con Herrera et al. (2018), la mayoría de ciber atacantes cursan el segundo y tercer año de secundaria, a diferencia de la mayoría de sus víctimas, que cursan el primer y segundo año. Adicionalmente, la diferencia de edad, el contenido y el tiempo de navegación en internet también resultaron ser factores a tener en cuenta a la hora de identificar a una persona agresora o víctima. Según Martin-Criado et al. (2021), las plataformas y actividades de mayor visita para los menores de edad se centran en las redes sociales y plataformas de video. Su alto consumo hace que los menores tiendan a recrear las situaciones, comportamientos y léxico que se exhiben en dichas plataformas, lo que incrementa la probabilidad de que un menor pueda recrear comportamientos violentos en su entorno.

Por otro lado, la minería de datos es una técnica que se ha popularizado en los últimos años; permite analizar grandes conjuntos de datos con el fin de encontrar y esclarecer cualquier patrón o relación no evidente a simple vista. Entre sus tareas más conocidas se encuentran el análisis, síntesis y visualización de datos. Su finalidad es generar conocimiento sobre temas de interés, ayudar en la toma de decisiones empresariales u optimizar procesos industriales. Con la generación masiva de información actual, las fuentes de datos se pueden encontrar en línea, a través de gestores de bases de datos o se generan mediante encuestas, grupos focales y otras técnicas de recolección. Entre los sectores más interesados en la minería de datos se encuentra el sector empresarial, que aprovecha este proceso para detectar anomalías, fraudes o mejorar sus estrategias comerciales. Existen metodologías maduras que facilitan y estandarizan sus fases, como el descubrimiento de conocimientos en las bases de datos (KDD) con sus seis etapas: selección de datos, preprocesamiento, transformación, minería de datos, evaluación y finalmente interpretación (Schröer et al., 2021).

El preprocesamiento de datos es una tarea que consiste en la limpieza y eliminación del ruido de los datos para ajustarse a técnicas y modelos de minería de datos. Estas tareas mejoran la calidad, consistencia y confiabilidad de los datos al corregir, reemplazar y eliminar datos incorrectos, irrelevantes o redundantes. La transformación, por otra parte, incluye la selección de atributos relevantes en el estudio, así como técnicas de transformación como la discretización, normalización y aumento o reducción de dimensionalidad. Adicionalmente, en caso de tratarse

de un tipo de aprendizaje supervisado, se debe establecer el atributo de salida o etiqueta (Castro R et al., 2018).

El aprendizaje supervisado se utiliza en situaciones específicas en donde se quiere inferir conocimiento a partir de datos etiquetados, es decir, que se conoce su variable a predecir o clasificar. En esta técnica se utilizan los datos etiquetados para entrenar el modelo, y se evalúa con datos no etiquetados para medir su rendimiento. Uno de los modelos utilizados en este estudio y más populares de aprendizaje supervisado es el árbol de decisión, un modelo fácil de implementar e interpretar cuando se configura adecuadamente. Su funcionamiento se basa en la estructura de un árbol, y representa a las características o atributos seleccionados mediante nodos, y a su vez, ramifica las opciones de cada una (Ramirez y Ccallohuari, 2020).

Por otra parte, el aprendizaje no supervisado no dispone de una variable de salida preestablecida, sino que trata de encontrar patrones o relaciones en el conjunto de datos que permitan clasificar, categorizar o etiquetar los registros. Una de las técnicas más utilizadas en el aprendizaje no supervisado y por ende en este estudio es el agrupamiento. Esta técnica trata de generar grupos con registros similares y separar los grupos con características ajenas. Su funcionamiento se basa en el uso de métricas de distancia o similitud, como la euclidiana, coseno, manhattan, entre otras (Roux, 2018). Existen variaciones de esta técnica con diferentes algoritmos internos y la elección de alguna dependerá del tipo de datos y el problema que se quiera afrontar (Bracco, 2018).

Materiales y Métodos

Los materiales necesarios para la aplicación de técnicas y algoritmos de minería de datos en este estudio se describen a continuación:

- Plataforma de minería de datos RapidMiner.
- Conjunto de datos de encuestas sobre experiencias con el acoso y ciberacoso en estudiantes universitarios.

RapidMiner es un software libre y gratuito (en su versión básica) de análisis y minería de datos que cuenta con una interfaz gráfica amigable hacia el usuario y un flujo de trabajo basado en operadores y conexiones, logrando mayor productividad en menos tiempo. Además, cuenta con más de 1,500 operadores que van desde tareas de preprocesamiento hasta técnicas avanzadas de modelado y visualización de datos (Kathuria et al., 2021). Asimismo, es compatible con los formatos de archivos que albergan los conjuntos de datos más utilizados como archivos de Excel. Este tipo de archivo contiene registros de datos de una encuesta realizada a estudiantes universitarios con el fin de recolectar información demográfica, estudiantil y experiencias con el acoso y el ciberacoso. El conjunto de datos cuenta con 702 registros y un total de 55 atributos o columnas.

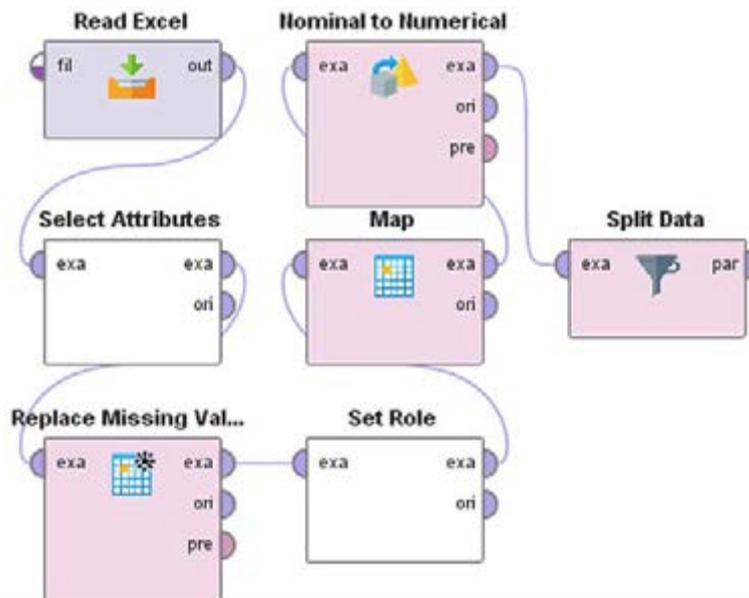
Para aplicar con éxito los procesos y algoritmos de minería de datos en cualquier entorno, es necesario aplicar la metodología KDD (Gupta y Chandra, 2020). Por lo tanto, es imprescindible una fase previa de conocimiento, familiarización y preprocesamiento de los datos, para comprender los tipos de datos, su extensión, longitud, atributos, inconsistencias y el atributo o etiqueta de salida en caso de tratarse de un aprendizaje supervisado (Ruiz-Chávez et al., 2018). La etiqueta es un valor que clasifica o categoriza al registro; en este contexto, un atributo que indica si el encuestado ha sufrido de acoso o ciberacoso. En esta fase se aplicaron los siguientes procesos:

- *Selección de atributos o variables:* se seleccionaron los atributos más relevantes de la encuesta realizada para el análisis y la sintetización textual del acoso y el ciberacoso. Para ello, se analizaron 18 atributos del cuerpo de la encuesta que proporcionan información necesaria para los procesos de agrupamiento y clasificación.
- *Limpieza o preprocesamiento de los datos:* para asegurar la calidad del proceso de minería de datos, los registros con valores nulos fueron identificados e imputados. Este proceso implica reemplazar los valores faltantes por otro valor, variable o estadístico.
- *Transformación de los datos:* muchos algoritmos y modelos de minería de datos requieren ciertos tipos de valores para funcionar correctamente o dar mejores resultados. Algunas de las técnicas de transformación de datos incluyen la discretización, codificación de valores y asignación de nuevos valores. Otra parte fundamental en este proceso es la división del conjunto de datos para entrenamiento y pruebas. En esta técnica es frecuente dividir los datos en 80% para entrenamiento y 20% para pruebas. Esta técnica garantiza la transparencia del modelo, debido a que será evaluado con datos desconocidos.

En RapidMiner, los operadores permiten al usuario llevar a cabo los procesos existentes en el programa, siempre y cuando se establezcan las conexiones correctas, los formatos y los tipos de datos que requieren. Los operadores utilizados para llevar a cabo dichos procesos y generar la síntesis textual se presentan en la Figura 1.

Figura 1

Esquema de operadores para preprocesamiento de datos

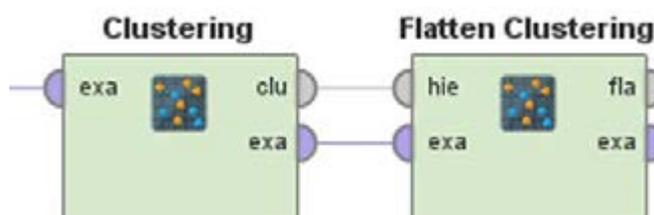


A partir de ahora se pueden aplicar modelos de clasificación o predicción, técnicas de agrupamiento y evaluación. En este estudio se utilizaron dos modelos de árboles de decisión, un modelo estándar, aplicado a partir de los 18 atributos seleccionados, y otro modelo generado a partir de un agrupamiento previo, etiquetando a cada grupo (clúster). En ambos modelos se utilizó el mismo proceso de síntesis previa, donde fueron evaluados y comparados en su tarea de clasificación. El agrupamiento jerárquico aglomerativo (HAC), árboles de decisión y métricas de evaluación fueron los modelos y algoritmos utilizados. Otros modelos, métricas de evaluación y procesos de optimización quedan fuera del alcance de este estudio.

Implementación de la técnica de agrupamiento

Las técnicas de agrupamiento se utilizan con frecuencia en la minería de datos para identificar grupos naturales en los datos a partir de su similitud (Mamani Rodríguez et al., 2017). Existen varios algoritmos de agrupamiento, pero para este estudio se utilizó el HAC debido a su estructura y simplicidad. Este agrupamiento jerárquico considera que cada punto de datos en un grupo individual, luego, en un algoritmo aglomerativo de “abajo hacia arriba” agrupa los puntos hasta formar grupos más grandes mientras sube en su jerarquía (Sharma et al., 2019). Las variables que se utilizaron para implementar la técnica de agrupamiento fueron los 18 atributos seleccionados que componen el cuerpo de la encuesta realizada a los estudiantes universitarios, debido a ser los más relevantes para realizar la síntesis textual. En conjunto, se utilizó un operador adicional que reduce en una sola jerarquía los grupos que se quieren generar, en este caso cuatro grupos. Los operadores que permiten implementar esta técnica se muestran en la Figura 2.

Figura 2
Operadores de agrupamiento



Los parámetros de configuración del operador de agrupamiento se modificaron para trabajar con los datos procesados. El modo o criterio que indica el tipo de enlazar cada punto en el agrupamiento, el tipo de medida indica la medida que se utilizará para medir la distancia entre los puntos, puede ser nominal, numérica o mixta, y a su vez la medida a utilizar. Este parámetro depende del anterior, ya que despliega distintas medidas en función de su tipo. Los parámetros establecidos en RapidMiner se presentan en la Tabla 1.

Tabla 1
Configuración de parámetros del agrupamiento

MODO	ENLACE PROMEDIO
Tipo de medida	Numérica
Medida numérica.	Coefficiente de similitud de Sorensen-Dice

Implementación de árboles de decisión

Los árboles de decisión son un modelo de aprendizaje automático supervisado. Su estructura se compone de nodos, que representan las características o atributos de los datos de entrada, y ramas, que representan las posibilidades de esos atributos (Fletcher e Islam, 2020). RapidMiner ofrece la posibilidad de generar árboles de decisión y modificar algunos de sus parámetros para optimizar su rendimiento.

Asimismo, como en la técnica anterior se utilizaron los 18 atributos preprocesados con la finalidad de obtener resultados más eficientes. Los parámetros que fueron modificados corresponden al criterio, que especifica la forma de selección y división de las ramas, así como la profundidad máxima, que limita el número de ramificaciones. Los valores modificados y el resto de los parámetros se evidencian en la Tabla 2.

Tabla 2
Configuración de parámetros del árbol de decisión

CRITERIO	RELACIÓN DE GANANCIA (GAIN RATIO)
Profundidad máxima	5
Confianza	0.10
Ganancia mínima	0.01
Tamaño mínimo de hojas	2

Evaluación de modelos

Para validar la eficacia de los modelos generados, se utilizaron operadores que permiten evaluar el rendimiento de los modelos a través de varios criterios como la precisión, la exactitud (Accuracy) y la recuperación (Recall). En RapidMiner, previo a la evaluación de los modelos, es necesario un operador que aplique el modelo con los datos de entrenamiento y prueba. Posteriormente, se conecta el operador que mide el rendimiento del modelo. Ambos operadores no disponen de parámetros de configuración.

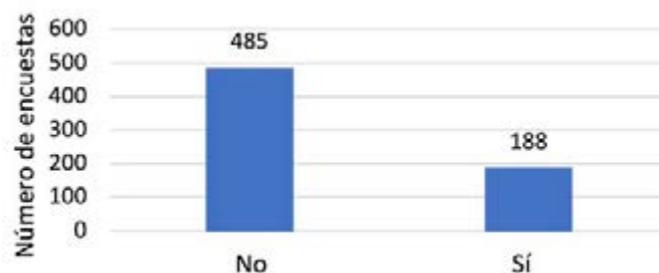
Finalmente, el rendimiento de los modelos fue evaluado a través de la matriz de confusión generada. De esta manera, fue posible comparar su rendimiento, así como los aciertos y errores en la clasificación de cada modelo, evidenciando cuál presenta las mejores prestaciones. Por consiguiente, los métodos descritos en esta sección son exclusivos de la tarea de síntesis textual sobre el acoso y el ciberacoso.

Resultados y Discusión

Los procesos y técnicas de preparación de los datos resultaron en la síntesis textual que se describe a continuación:

- Selección de atributos o variables: se seleccionaron 18 atributos del cuerpo de la encuesta que incluyen datos ordinales (ítems tipo Likert), un identificador (número de encuesta) y un atributo de salida. Este atributo etiqueta al estudiante que ha sufrido de acoso o ciberacoso mediante dos valores (Sí y No).
- Limpieza de datos: al no presentarse más problemas de limpieza de datos, los valores nulos fueron imputados al valor de la media. El número de registros luego de la limpieza se redujo a 673. La Figura 3 muestra la proporción de estudiantes que han sido víctimas de acoso o ciberacoso en la secundaria.

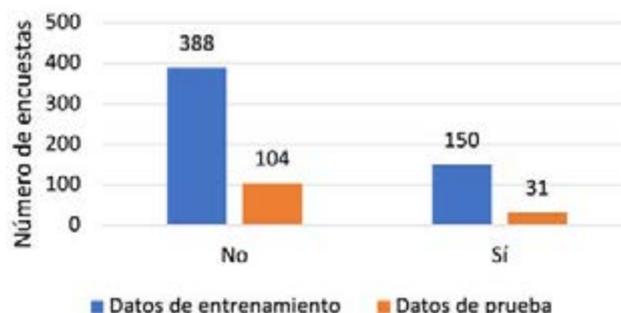
Figura 3
Frecuencia de experiencias con el acoso



- Transformación de los datos: los valores de los atributos o columnas seleccionadas (4. Muchas veces; 3. Bastantes veces; 2. Algunas veces y 1. Nunca) se reemplazaron por dos variables nominales (Sí y No). La división de datos se realizó con un porcentaje de 80% (538 registros) para entrenamiento y 20% (135 registros) para pruebas, como se muestra en la Figura 4.

Figura 4

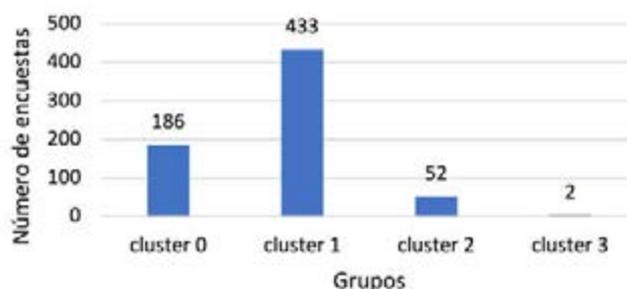
Frecuencia de experiencias en datos de entrenamiento y prueba



- Para el agrupamiento, se transformaron los valores de los registros a un valor numérico (0 y 1), aumentando una columna por cada valor del registro. Esta codificación permitió que el método de agrupamiento jerárquico aglomerativo pudiera procesar y realizar los cálculos numéricos necesarios para el agrupamiento. En la Figura 5 se muestra los cuatro grupos generados en RapidMiner.

Figura 5

Grupos generados



En RapidMiner es posible visualizar los modelos de árboles de decisión a través de gráficos que representan los nodos y ramificaciones del modelo. Aunque en algunas ocasiones, los modelos resultan extensos debido a la cantidad de atributos o su configuración, como ha ocurrido en este caso, por lo que una representación gráfica de los árboles no es adecuada en esta ocasión.

Sin embargo, la matriz de confusión expone el rendimiento del modelo a través de varias métricas, como la precisión para las clases, que indica la proporción de verdaderos positivos sobre el total de calificaciones positivas. Y la métrica de recuperación (Recall) de cada clase, que indica la proporción de verdaderos positivos que se han clasificado correctamente. Esto es realizado a través de la ecuación 1 y 2:

$$precisión = \frac{TP}{TP + FP} \quad (1)$$

$$\text{recuperación} = \frac{TP}{TP + FN} \quad (2)$$

Donde TP es verdadero positivo, FP es falso positivo y FN es falso negativo.

El modelo de clasificación sin agrupamiento previo presentó una exactitud del 78.52%, y el modelo de clasificación con agrupamiento presentó una exactitud del 86.57%. En las Tabla 3 y la Tabla 4 exponen la matriz de confusión de ambos modelos, así como los valores de precisión y el porcentaje de recuperación.

Tabla 3
Matriz de confusión del modelo sin agrupamiento

	SI	NO	PRECISIÓN DE LA CLASE
PRED. NO	92	24	79.31%
PRED. SÍ	5	14	73.68%
RECUPERACIÓN DE LA CLASE	94.85%	36.84%	

La precisión de ambas clases de acuerdo con la Tabla 3 es superior al 70%. Por lo tanto, el modelo puede clasificar con precisión al menos esa proporción de registros de estudiantes que sufren de acoso o ciberacoso. Sin embargo, la recuperación de la clase Sí presenta algunas dificultades al modelo, acertando en 14 ocasiones de un total de 38 registros pertenecientes a la clase Sí, dando como resultado una recuperación del 36.84%. Por el contrario, la clase No, que presenta una proporción del 94.85%, acertó en 92 ocasiones de un total de 97 registros de esta clase.

De acuerdo con la Tabla 4, el modelo con agrupamiento previo pudo clasificar a tres de los cuatro grupos formados con una precisión de al menos 75%. Sin embargo, al dividir los datos aleatoriamente en un grupo de entrenamiento y de prueba, no quedaron registros existentes del grupo tres en los datos de prueba. En cuanto a la recuperación, el grupo dos presentó una proporción del 100%, mientras que el grupo uno fue del 97.70%. Lo que acertó en clasificar 85 casos de 87 y finalmente el grupo cero cuya una proporción fue del 56.76%, llegó a clasificar 21 casos correctos de un total de 37 casos pertenecientes a ese grupo.

Tabla 4
Matriz de confusión del modelo con agrupamiento

	CLUSTER 0	CLUSTER 1	CLUSTER 2	CLUSTER 3	PRECISIÓN DE LA CLASE
PRED. CLUSTER 0	21	2	0	0	91.30%
PRED. CLUSTER 1	13	85	0	0	86.73%
PRED. CLUSTER 2	3	0	10	0	76.92%
PRED. CLUSTER 3	0	0	0	0	0.00%
RECUPERACIÓN DE LA CLASE	56.76%	97.70%	100.00%	0.00%	

La síntesis textual facilitó la generación de modelos de clasificación; ambos modelos presentaron una exactitud por encima del 75%. Sin embargo, los resultados evidencian una

mayor exactitud al implementar el modelo con agrupamiento, lo que indica que el proceso de síntesis tuvo una influencia positiva en su rendimiento al predecir los casos correctos de las personas que han sufrido de acoso y ciberacoso. Aun así existen otros factores que pueden alterar los resultados de un modelo de minería de datos, como las tareas de preprocesamiento de datos, la calidad de los datos y el balance de las clases o etiquetas.

Conclusiones

La minería de datos se encuentra en un proceso continuo de avance y transformación, al igual que sus herramientas de software. La generación masiva de información permite la creación de nuevos algoritmos y procesos que permiten analizar y generar conocimiento de cualquier tipo de información, incluso de problemas sociales tan delicados como el acoso y el ciberacoso. Cada día, más jóvenes y adolescentes sufren estos tipos de violencia en sus actividades cotidianas; esto se evidencia en el número de estudiantes universitarios que respondieron afirmativamente a la encuesta sobre sus experiencias con el acoso y el ciberacoso.

Por lo tanto, en este estudio se utilizaron técnicas y algoritmos de minería de datos como árbol de decisión y de agrupamiento a las encuestas aplicadas a los estudiantes universitarios para generar una síntesis textual que permita generar conocimiento sobre el acoso y el ciberacoso. De este modo, al aplicar el modelo de aprendizaje automático supervisado denominado árbol de decisión con los datos obtenidos de las encuestas presentó una exactitud del 78.52%. Sin embargo, al evaluar con los mismos datos sobre el acoso y el ciberacoso, el modelo de clasificación con agrupamiento alcanzó una exactitud del 86.57% al estimar cuando se presentan estos problemas sociales a través de la encuesta realizada. No obstante, es imprescindible realizar tareas de preprocesamiento de datos para lograr un buen desempeño de los modelos, debido a que pueden alterar los resultados del modelo que se está utilizando.

Como trabajos futuros, se plantea el uso de otros modelos de aprendizaje automático y aprendizaje profundo como las redes neuronales; así como otras técnicas de evaluación de modelos como la validación cruzada que pueden ser de utilidad para la generación de conocimiento sobre el acoso y el ciberacoso.

Reconocimientos

Los autores desean agradecer al Vicerrectorado de Investigaciones de la Universidad del Azuay por el apoyo financiero y académico, así como a todo el personal de la escuela de Ingeniería de Ciencias de la Computación, y el Laboratorio de Investigación y Desarrollo en Informática (LIDI).

Referencias

- Arce-Ruelas, K. I., Álvarez-Xochihua, O., Pelegrín, L., Cardoza-Avendaño, L., y González-Fraga, J. A. (2022). Automatic Cyberbullying Detection: A Mexican Case in High School and Higher Education Students. *IEEE Latin America Transactions*, 20(5).
- Banerjee, V., Telavane, J., Gaikwad, P., y Vartak, P. (2019). Detection of Cyberbullying Using Deep Neural Network. *International Conference on Advanced Computing and Communication Systems (ICACCS)*.
- Bozyigit, A., Bilimleri, B., Dokuz, B., İzmir, E. Ü., Bilgisayar, S. U., Bölümü, M., Eylül, D., İzmir, Ü., Bilgisayar, E. N., y Bölümü, B. (2019). Cyberbullying Detection by Using Artificial Neural Network Models. *International Conference on Computer Science and Engineering (UBMK)*.

- Bracco, A. (2018). *Normalización de Texto en Español de Argentina* (pp. 1–68).
- Castilla, O. M. N. (2021). Cyberbullying: El acoso escolar en el ciberespacio e implicancias psicológicas. *Hamut'ay*, 8(1), 67–74.
- Castro R, L. F., Espitia P, E., y Montilla, A. F. (2018). Applying CRISP-DM in a KDD process for the analysis of student attrition. *Advances in Computing: 13th Colombian Conference, CCC 2018, Cartagena, Colombia, September 26–28, 2018, Proceedings 13*, 386–401.
- Cornel, J. A., Pablo, C. C., Marzan, J. A., Mercado, J. V., Fabito, B., Rodriguez, R., Octaviano, M., Oco, N., y la Cruz, A. De. (2019). Cyberbullying Detection for Online Games Chat Logs using Deep Learning. *International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment and Management (HNICEM)*.
- Dalvi, R. R., Chavan, S. B., y Halbe, A. (2020). Detecting a Twitter Cyberbullying using Machine Learning. *International Conference on Intelligent Computing and Control Systems (ICICCS)*.
- Fletcher, S., y Islam, Md. Z. (2020). Decision Tree Classification with Differential Privacy. *ACM Computing Surveys*, 52(4), 1–33. <https://doi.org/10.1145/3337064>
- Gupta, M. K., y Chandra, P. (2020). A comprehensive survey of data mining. *International Journal of Information Technology*, 12(4), 1243–1257. <https://doi.org/10.1007/s41870-020-00427-7>
- Han, J., Pei, J., y Tong, H. (2022). *Data mining: concepts and techniques*. Morgan kaufmann.
- Herrera, C. R. M., Ríos, S. P., y Noboa, I. R. (2018). Indicadores de violencia relacionados con el cyberbullying en adolescentes del Ecuador. *Pensando Psicología*, 14(24).
- Jorgensen, S. L., y Siegel, P. B. (2019). *Social Protection in an Era of Increasing Uncertainty and Disruption*. World Bank, Washington, DC. <https://doi.org/10.1596/31812>
- Kathuria, A., Gupta, A., y Singla, R. K. (2021). *A Review of Tools and Techniques for Preprocessing of Textual Data* (pp. 407–422). https://doi.org/10.1007/978-981-15-6876-3_31
- Li, C., Wang, P., Martin-Moratinos, M., Bella-Fernández, M., y Blasco-Fontecilla, H. (2022). Traditional bullying and cyberbullying in the digital age and its associated mental health problems in children and adolescents: a meta-analysis. *European Child y Adolescent Psychiatry*, 1–15.
- Mahat, M. (2021). Detecting Cyberbullying across Multiple Social Media Platforms Using Deep Learning. 2021 *International Conference on Advance Computing and Innovative Technologies in Engineering, ICACITE 2021*, 299–301. <https://doi.org/10.1109/ICACITE51222.2021.9404736>
- Martin-Criado, J. M., Casas, J. A., Ortega-Ruiz, R., y Rey, R. Del. (2021). Parental supervision and victims of cyberbullying: Influence of the use of social networks and online extimacy. *Revista de Psicodidactica*, 26(2), 161–168. <https://doi.org/10.1016/j.psicod.2020.12.005>
- Mollo, J. P., Larrain, E., y Landazabal, M. G. (2018). Prevalencia de bullying y cyberbullying en Latinoamérica: una revisión. *Revista Iberoamericana de Psicología: Ciencia y Tecnología*, 11(3), 1–18.
- Namane, K. C., y Kyobe, M. (2017). Examining the evolution of Mobile Bully - Victims across different schools located in low to high safety risk areas in Cape Town, South Africa. *2017 Conference on information Communication Technology and Society*.
- Nielsen, M. B., y Einarsen, S. V. (2018). What we know, what we do not know, and what we should and could have known about workplace bullying: An overview of the literature and agenda for future research. *Aggression and Violent Behavior*, 42, 71–83. <https://doi.org/10.1016/j.avb.2018.06.007>
- Rahman, S., Talukder, K. H., y Mithila, S. K. (2021). An Empirical Study to Detect Cyberbullying with TF-IDF and Machine Learning Algorithms. *Proceedings of International Conference on Electronics, Communications and Information Technology, ICECIT 2021*. <https://doi.org/10.1109/ICECIT54077.2021.9641251>

- Ramirez, A. J. B., y Ccallohuari, H. A. M. (2020). *Modelo de aprendizaje supervisado para pronóstico de la deserción de estudiantes de la Facultad de Ingeniería y Arquitectura de la Universidad Peruana Unión - Lima*.
- Roux, M. (2018). A comparative study of divisive and agglomerative hierarchical clustering algorithms. *Journal of Classification*, 35, 345–366.
- Ruiz-Chavez, Z., Salvador-Meneses, J., y Garcia-Rodriguez, J. (2018). *Machine Learning Methods Based Preprocessing to Improve Categorical Data Classification* (pp. 297–304). https://doi.org/10.1007/978-3-030-03493-1_32
- Schröer, C., Kruse, F., y Gómez, J. M. (2021). A systematic literature review on applying CRISP-DM process model. *Procedia Computer Science*, 181, 526–534.
- Shaikh, F. B., Rehman, M., y Amin, A. (2020). Cyberbullying: A Systematic Literature Review to Identify the Factors Impelling University Students towards Cyberbullying. *IEEE Access*, 8, 148031–148051. <https://doi.org/10.1109/ACCESS.2020.3015669>
- Sharma, S., Batra, N., y others. (2019). Comparative study of single linkage, complete linkage, and ward method of agglomerative clustering. *2019 international conference on machine learning, big data, cloud and parallel computing (COMITCon)*, 568–573.
- Tapia, F., Aguinaga, C., y Luje, R. (2018). Detection of Behavior Patterns through Social Networks like Twitter, using Data Mining techniques as a method to detect Cyberbullying. *2018 7th International Conference On Software Process Improvement (CIMPS)*, 111–118.

TICeC2023

Información para Autores



TRAVEL is to take a journey or to have an adventure to somewhere by bicycle, train, airplane, car, motorcycle, or boat. It could be an exploration to somewhere new planned or unplanned to meet new people, new things and new places. There are different types of adventures waiting for you to explore.

There are lots of places to explore. Places could be urban or suburban. Some people love to be with nature to free their minds and refresh their souls, but some like to be in the city. You will get lots of benefits such as exploring new culture.

Evaluación por pares

La *Revista Tecnológica Espol - RTE* es una revista arbitrada que se rige por el sistema doble par anónimo. Los artículos enviados por los autores son evaluados en previamente por el Comité de Redacción para comprobar si se ajustan a las normas de edición y a las políticas temáticas de la revista. Cuando el artículo pasa ese primer filtro es enviado a dos evaluadores externos expertos en la temática abordada por el autor. Para cumplir y defender la ética de la investigación, estos evaluadores desconocen el nombre de el/los autores/as y la identificación de la/s institución/es a la que pertenece el artículo, encargándoseles dictaminar si responde a los intereses científicos de la revista y si procede su publicación. En la valoración final, los revisores deciden entre las siguientes opciones: publicable, publicable con modificaciones menores, publicable con modificaciones mayores o no publicable. En el caso de que haya disparidad de opinión entre revisores del Comité Científico, se someterá al juicio de un tercer experto, que dirimirá en conflicto de pareceres.

En el siguiente link [Guía de revisores](#), se detalla paso a paso el proceso de evaluación a seguir por los pares ciegos asignados

Las responsabilidades específicas de los revisores se encuentran declaradas en el link [Código de ética y buenas prácticas de publicación](#).

Derechos de autor (Copyright)

Los originales publicados, en las ediciones impresa y electrónica, de la *Revista Tecnológica Espol - RTE*, bajo derechos de primera publicación, son propiedad de la Escuela Superior Politécnica del Litoral (ESPOL), Guayaquil, República del Ecuador, siendo absolutamente necesario citar la procedencia en cualquier reproducción parcial o total de los contenidos (textos o imágenes) publicados. RTE proporciona un acceso abierto e inmediato a su contenido, pues creemos firmemente en el acceso público al conocimiento, lo cual no obsta para que la cita de la fuente sea obligatoria para todo aquél que desee reproducir contenidos de esta revista.

De igual modo, la propiedad intelectual de los artículos o textos publicados en la revista RTE pertenece al/la/los/las autor/a/es/as, quienes conservan sus derechos de autor y dominio irrestricto de su obra.

Esta circunstancia ha de hacerse constar expresamente de esta forma cuando sea necesario.

Todo el contenido de RTE mantiene una licencia de contenidos digitales otorgada por Creative Commons.



Esta obra está bajo una Licencia Creative Commons
Atribución-NoComercial 4.0 International.

Directrices para autores

Los textos postulados deben:

1. Corresponder a las categorías universalmente aceptadas como producto de investigación.
2. Ser originales e inéditos.
3. Sus contenidos responden a criterios de precisión, claridad y brevedad.

Se clasifican en:

3.1. Artículos. En esta sección se publican:

- 3.1.1. Artículos de investigación científica o tecnológica: presenta de manera detallada los resultados originales de proyectos terminados de investigación. La estructura generalmente utilizada contiene cuatro aportes importantes: introducción, metodología, resultados y conclusiones.
- 3.1.2. Artículo de reflexión o ensayo: presenta resultados de investigación terminada desde una perspectiva analítica, interpretativa o crítica del autor, sobre un tema específico recurriendo a fuentes originales.
- 3.1.3. Artículo de revisión: resultado de una investigación terminada donde se analizan, sistematizan e integran los resultados de investigaciones, publicadas o no, ya sea en el campo científico, artístico o artístico tecnológico, con el fin de dar cuenta de los avances y las tendencias de desarrollo.

Próximos Volúmenes

PROGRAMACIÓN

Vol. 35, Núm. 3

Volumen Abierto Semestral

- i. Convocatoria abierta: hasta el 15 de julio de 2023
- ii. Artículos seleccionados para publicación: hasta el 19 de agosto de 2023
- iii. Publicación del volumen: 30 de diciembre de 2023

Sección Especial EDUCAMP 2023

- i. Convocatoria abierta: hasta el 30 de septiembre de 2023
- ii. Artículos seleccionados para publicación: hasta el 30 de noviembre de 2023
- iii. Publicación del volumen: 30 de diciembre de 2023

Más información: <http://www.rte.espol.edu.ec/public/site/educamp2023-es.pdf>

Próximos Volúmenes

PROGRAMACIÓN

Vol. 36, Núm. 1

Volumen Abierto Semestral

- i. Convocatoria abierta: hasta el 30 de abril de 2024*
- ii. Artículos seleccionados para publicación: hasta el 15 de mayo de 2024*
- iii. Publicación del volumen: 30 de junio de 2024*

PROGRAMACIÓN

Vol. 36, Núm. 2

Volumen Abierto Semestral

- i. Convocatoria abierta: hasta el 30 de septiembre de 2024*
- ii. Artículos seleccionados para publicación: hasta el 30 de noviembre de 2024*
- iii. Publicación del volumen: 30 de diciembre de 2024*

PROGRAMACIÓN

Vol. 37, Núm. 1

Volumen Abierto Semestral

- i. Convocatoria abierta: hasta el 30 de abril de 2025*
- ii. Artículos seleccionados para publicación: hasta el 15 de mayo de 2025*
- iii. Publicación del volumen: 30 de junio de 2025*

NOTE: For interest in publications of special issues, universities and organizations related to research are invited to contact us at nsolorza@espol.edu.ec

NOTA: Para interés en publicaciones de volúmenes especiales, se invita a universidades y organizaciones relacionadas con investigación a contactarnos a nsolorza@espol.edu.ec





espol

cedia



rte.espol.edu.ec