

Evaluación de sitios turísticos mediante análisis de sentimientos de comentarios emitidos por usuarios en redes sociales

Evaluation of tourist sites through sentiment analysis of comments issued by users on social networks

Nelson Herrera Herrera^{1, 2} <https://orcid.org/0000-0002-5781-6444>,
Nelson Salgado Reyes^{1, 3} <https://orcid.org/0000-0001-8908-7613>

¹*Instituto Superior Tecnológico Japón, Quito, Ecuador*
nherrera@itsjapon.edu.ec, nsalgado@itsjapon.edu.ec

²*Universidad UTE, Quito, Ecuador*
nelson.herrera@ute.edu.ec

³*Pontificia Universidad Católica, Quito, Ecuador*
nesalgado@puce.edu.ec



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

Enviado: 2022/02/28

Aceptado: 2022/05/09

Publicado: 2022/06/30

Resumen

Esta investigación tiene como objetivo presentar la utilidad del análisis de sentimientos en los comentarios emitidos por usuarios de servicios turísticos en redes sociales (Twitter y Trip Advisor), que permita calificar el nivel de dichos servicios. La ciudad de Quito-Ecuador es considerada como caso de estudio para este trabajo. La investigación se muestra en función del desarrollo de un sistema informático utilizando herramientas Big Data (Python, Node.js, Mongo DB), que permiten recolectar, almacenar y procesar gran cantidad de información. Las librerías del lenguaje de programación Node.js: Puppeteer y Sentiment, permiten obtener los comentarios de las redes sociales Twitter y Trip Advisor, y determinar un puntaje sobre el destino turístico evaluado. Entre los aspectos novedosos de la investigación está el uso de la red social Twitter como fuente de origen de datos, también la utilización de técnicas de web scraping del sitio de Trip Advisor. Para este estudio se emplea la interfaz de programación de aplicaciones de Twitter (Application Programming Interface, API), la cual permite tomar datos de esta red social en tiempo real que facilita evaluar los servicios turísticos ofertados. Como resultado se puede determinar que la herramienta permite generar conocimiento que ayuda a

Sumario: Introducción, Trabajos Relacionados, Materiales y Métodos, Metodología, Selección de las herramientas, Resultados y Discusión y Conclusiones.

Como citar: Herrera, N. & Salgado, N. (2022). Evaluación de sitios turísticos mediante análisis de sentimientos de comentarios emitidos por usuarios en redes sociales. *Revista Tecnológica - Espol*, 34(2), 125-139. <http://www.rte.espol.edu.ec/index.php/tecnologica/article/view/921>

tomar decisiones, relacionadas con el nivel de calidad de las prestaciones recibidas en los sitios visitados.

Palabras clave: Big Data, redes sociales, turismo, análisis de sentimientos.

Abstract

This research aims to present the usefulness of sentiment analysis in the comments issued by users of tourism services on social networks (Twitter and Trip Advisor), which allows rating the level of such services. This case study is set in the city of Quito-Ecuador. The research develops a computer system using Big Data tools (Python, Node.js, Mongo DB) to collect, store, and process large amounts of information. The Node.js programming language libraries, Puppeteer and Sentiment, make it possible to obtain comments from the social networks Twitter and Trip Advisor and determine a score for the observed tourist destination. Among the novel aspects of this research is the use of the social network Twitter as a source of data origin and web scraping techniques from the Trip Advisor site. This study uses the Twitter Application Programming Interface (API) to obtain data from this social network in real-time, facilitating the evaluation of the tourist services. Results determine that this tool allows the generation of knowledge to decide the quality level of the services received in the visited sites.

Keywords: Social Networks, Tourism, Sentiment Analysis, Big Data.

Introducción

En la economía de cualquier país del mundo, el turismo constituye un factor importante que origina múltiples oportunidades de empleo y permite mejorar el nivel de vida de los habitantes. Ecuador consiente de esta realidad desarrolla un plan estratégico institucional, que establece lineamientos que ubica al turismo entre las primeras fuentes de ingreso del país (Ministerio de Turismo del Ecuador, 2019).

Mejorar la calidad de los servicios turísticos depende, entre otros aspectos, del criterio de los usuarios. Sin embargo, este criterio muchas veces se lo recibe en función de encuestas, entrevistas informales y otros mecanismos, que no ofrecen información útil para poder tomar decisiones en tiempo real. Esto constituye un problema a la hora de determinar acciones de rediseño y mejora de los servicios ofertados por los destinos turísticos.

Existen algunas páginas web o redes sociales que recomiendan destinos turísticos. Se basan en una calificación de los usuarios, pero no en una evaluación basada totalmente en el comentario de los consumidores de un servicio, sino al análisis de sentimientos en los comentarios emitidos por los mismos a través de las redes sociales. Al no usar el gran volumen de comentarios generados por varios usuarios, se pierde información muy valiosa que puede ayudar a tomar las mejores decisiones o conocer si un destino turístico puede mejorar en algún área específica para atraer turistas.

En función a la problemática planteada se requiere un sistema que permita: recolectar, almacenar, procesar y visualizar información semi estructurada y no estructurada en tiempo real; que facilite analizar la calidad de los servicios turísticos. El objetivo de este trabajo consiste en desarrollar un sistema para la evaluación de destinos turísticos, mediante el análisis de sentimientos en los comentarios emitidos por usuarios de servicios turísticos, provenientes de redes sociales. Las fuentes de datos que la aplicación utiliza son las redes sociales: Twitter y Trip Advisor. El análisis de sentimientos permite obtener un puntaje donde se determina la calidad de dichos servicios.

Twitter dispone de una Application Programming Interface (API) abierta que permite a los desarrolladores integrar esta red social como un servicio en sus aplicaciones web, de escritorio o móviles. Esta compañía se caracteriza por mantener poca o casi ninguna protección del contenido publicado por los usuarios en su red social, lo cual es una ventaja para el desarrollador, pero ha ocasionado inconvenientes para sus usuarios con referencia al tema de seguridad de datos (Rik & Tim de Boer, 2011).

Por otra parte, Trip Advisor es una herramienta utilizada mundialmente y desde 2018, se considera una red social ya que se puede seguir a otros usuarios y compartir opiniones sobre sitios turísticos que se encuentran alrededor del mundo. En Ecuador Trip Advisor ofrece una gama amplia de noticias vinculadas con la cultura, gastronomía, hotelería y turismo de las diferentes ciudades, centrandolo en las ciudades de Quito, Guayaquil, Cuenca y Manabí. En la actualidad, esta red social es el sitio más reconocido, utilizado y confiable para los viajeros que buscaban diferentes opiniones sobre destinos, hoteles, atracciones turísticas, entre otras. Estadísticas manifiestan que la página recibe alrededor de 280 millones de visitas mensuales al igual que millones de comentarios y calificaciones de diferentes atracciones turísticas (Escribano, 2017).

El artículo realiza una presentación técnica del sistema y sus componentes; sin embargo, la investigación efectuada incluye que todos los elementos deben ser considerados. De acuerdo a lo mencionado se realiza la fundamentación teórica que se basa en el estudio, aplicación de técnicas y estrategias utilizadas en el análisis de sentimientos.

El presente trabajo tiene el contexto de una investigación Big Data (Zicopoulos et al., 2011), debido a que los elementos tecnológicos se conjugan perfectamente:

Volumen: gran cantidad de datos son generados mediante la red social Twitter y la proporcionada por el sitio Trip Advisor.

Variedad: datos generados por la red social Twitter y el sitio Trip Advisor.

Velocidad: a diario se generan datos mediante la red social Twitter a una considerable velocidad, lo cual requiere que el procesamiento y análisis posterior se lo realice en tiempo real, que permite tomar decisiones adecuadas (Aguilar, 2013).

Veracidad: la información de la red social Twitter, debe ser filtrada para elegir, información útil para la investigación.

Trabajos Relacionados

En el ámbito mundial se han realizado varias investigaciones relacionadas con la utilización de herramientas tecnológicas en el sector turístico.

- Acceso a una descripción del perfil del turista, mediante el análisis de miles de datos generados a cada instante, el visitante deja su huella digital, lo cual permite conocer sus preferencias y gustos que ayudan a los sitios turísticos a ofertar productos y servicios personalizados logrando fidelizar y atraer nuevos clientes (Tello, 2019).
- Sistema para tomar decisiones, para cualquier tipo de organización, incluido el sector hotelero. Contar con información de comportamientos del consumidor, información de reservas, formas de pago entre otras características. Genera conocimiento que al momento de tomar decisiones podría representar una

herramienta de gran valor es, en el caso puntual, de los servicios de hospedaje (Padilla, 2019).

- Sistema para análisis de mercado mediante el uso de Redes Sociales. Para realizar estudios de mercado las empresas deben buscar alternativas para extraer información y posterior análisis de datos, que permita a las organizaciones conseguir una ventaja competitiva. El trabajo concluye que las 4 plataformas que mayor beneficio ofrecen para esta investigación son: Google Trends, Facebook, YouTube y Twitter (Fajardo et al., 2021).
- Turismo y Big Data. Analizar perfiles, etiquetas y ubicaciones, así como metadatos que ofrece la plataforma Instagram para realizar estudios relacionados con el sector turístico. También se plantea un análisis automatizado de las imágenes exportable a grandes conjuntos de datos que suministrarán información necesaria para conseguir un cuadro de mando integral propicio y útil en la toma de decisiones en el sector turístico (Martínez & Piñeiro, 2020).
- Plataforma para el Análisis de Mercado a través de datos de Redes Sociales. Este estudio muestra el beneficio de la utilización de redes sociales al momento de realizar estudios de mercado (Fajardo et al., 2021).

En este ámbito otra de las redes sociales potentes que admiten el análisis de datos en tiempo real, de los mensajes expuestos y poder realizar acciones de mejora a los servicios turísticos, es la red social Twitter. Un insumo adicional importante al momento de obtener comentarios de los servicios ofertados es el sitio Trip Advisor, que concentra opiniones de quienes reciben los servicios de hoteles, restaurantes y operadores turísticos. La utilización de estas dos fuentes de datos permite tener dos criterios previos que, finalmente mediante contraste, ayuda a la generación de información y conocimiento frente a la creación o rediseño de los servicios ofertados.

Materiales y Métodos

Herramientas utilizadas

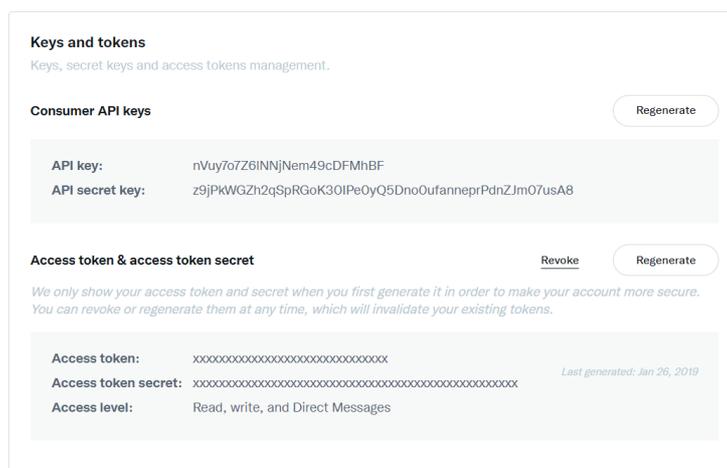
Para evaluar la oferta de los servicios de lugares turísticos, se considera la red social Twitter por su utilidad y relevancia. Por otra parte, se incluye al sitio web Trip Advisor debido a ser considerada una red social exclusiva para obtener información de lugares turísticos, que resulta valiosa al momento de realizar, análisis de sentimientos con los comentarios emitidos por los usuarios.

Datos de Twitter

La primera fuente de datos empleada para el estudio es Twitter, estos datos sirven para el análisis de la información rescatada de los usuarios de esta red social, para ello, se requiere tener una cuenta en dicha red; además debe estar activo el servicio de GPS en el dispositivo. Esta red social guarda comentarios de los clientes de los diversos sitios turísticos que están registrados, lo que permite tomar esta información para su posterior análisis para el efecto se usa la API de Twitter.

Para poder utilizar la API de Twitter se debe tener una cuenta de desarrollador en esta red social, que permite obtener un token o llave que se usa para tener acceso a la información pública de Twitter. Cada usuario tiene la decisión de otorgar permisos adicionales a todas las aplicaciones autorizadas a través de las opciones de su cuenta (Twitter, 2022).

Figura 1
Llaves y Tokens utilizados



Nota: Credenciales entregadas por la API de Twitter

Trip Advisor

Es un sitio web creado en 2000, que tiene como función compartir las opiniones y experiencias de profesionales sobre diferentes atracciones turísticas. El servicio inicial es para Estados Unidos. Ahora Trip Advisor se emplea mundialmente y, desde el 2018, se considera una red social que permite seguir a otros usuarios y compartir opiniones sobre sitios turísticos que se encuentran alrededor del mundo (Arias, 2017).

La técnica Web Scraping se utiliza para recolectar datos de sitios web. Dichos datos resultan de gran importancia para determinar patrones que establezcan el comportamiento de los usuarios. En este estudio, dicha herramienta se aplica a la red social Trip Advisor para obtener comentarios de los sitios cuyos servicios van a ser evaluados.

Metodología

La revisión bibliográfica de casos de éxito de la utilización de redes sociales aplicadas al sector turístico, sirvió para analizar y evaluar aplicaciones utilizadas en el ámbito mundial, relacionadas con el análisis de sentimientos en comentarios emitidos por usuarios en redes sociales.

Esto permitió respaldar el fundamento teórico y contenido metodológico de la presente investigación. En la investigación realizada por (Algecira, 2020) se establece la arquitectura base para proyectos de analítica de datos orientados al sector turístico.

Fase de generación

Al momento de considerar el diseño del sistema se tomaron en cuenta varios factores. El primer factor es determinar las redes sociales que se podrían usar, siendo la base para la investigación, debido a que una de las redes más importantes, Facebook, bloquea su acceso a la API para obtener comentarios de cualquier usuario (Krishna, 2018). Los comentarios de Twitter se enfocaron en la cuenta de un solo usuario, lo que quiere decir que, al realizar un análisis de sentimientos se obtuvieron comentarios que no podrían tener algún tipo de sentimiento, es decir neutros, porque se basan en la opinión de un solo usuario que coloca actualizaciones sobre un lugar. En otra instancia, Trip Advisor tiene un conjunto de comentarios basado en la opinión de varios usuarios, lo que implica que la probabilidad de que un comentario sea neutro en el análisis de sentimientos es muy baja.

Una vez determinada las fuentes de datos que fueron utilizadas en el proceso de análisis de sentimientos, se generaron datos que a continuación sirvieron como entrada, para el proceso primordial del sistema. En este contexto las fuentes que generan información son las redes sociales: Twitter y Trip Advisor. Las dos fuentes utilizadas proporcionaron datos de comentarios emitidos por usuarios de las redes sociales mencionadas anteriormente, cabe mencionar que los datos de Twitter son de tipo no estructurado¹ y los del sitio Trip Advisor de tipo semiestructurado².

Fase de Recolección

En este caso se utilizó la API de Twitter para recolectar mensajes de usuarios, en los cuales se pudo evidenciar la opinión de los ciudadanos, relacionados con la conformidad o no de los servicios ofertados en lugares turísticos.

El siguiente componente importante en la investigación fueron los datos generados por la red social Trip Advisor, la misma que registró información de reseñas y comentarios relacionados con el sitio visitado, los comentarios se los obtuvo mediante minería de opiniones. Para ello, se utilizó la librería Puppeteer de Node.js. Esta librería a la vez emplea el navegador Chromium, y mediante Web Scraping permite navegar automáticamente en cualquier página web y acceder al texto de la misma, en este caso, a comentarios mediante etiquetas internas del código HTML de la página.

Fase de Integración

Para agrupar los datos provenientes de las dos fuentes de datos utilizadas en la investigación se utilizó la base de datos NoSql MongoDB, debido que esta base de datos tiene fácil integración con el lenguaje Node.js y existen algunas librerías que la ayudan a tener mayor funcionalidad. Además, considerando que el funcionamiento de la aplicación es en tiempo real se determinó un repositorio que permitió al sistema manejar grandes volúmenes de información de manera eficaz y eficiente, la cantidad de comentarios a ser no pudo ser determinados pues dependen de varios factores y, en muchos casos, depende de que tan popular sería la red social, el número de visitas y los comentarios que las personas podrían publicar por día o por semana.

La Tabla 1 muestra el esquema general los campos en la tabla de la base de datos del sistema.

Como se puede observar la base de datos guardó: el nombre de la red social, el lugar, el comentario más actual, un puntaje positivo, negativo, el número de comentarios positivos y el número de comentarios negativos. Estos datos incluyen cuatro parámetros:

- ✓ Puntaje en función del precio
- ✓ Puntaje en función del servicio
- ✓ Puntaje en función del transporte
- ✓ Puntaje en función del hospedaje

Finalmente, el conjunto de tuits recolectados fue integrado con los datos que se obtuvieron de la red social Trip Advisor, para su posterior análisis, procesamiento y visualización de la información.

¹ Son datos binarios que no tienen estructura interna identificable.

² Se refieren a cualquier información que utilice un esquema de autodescripción: HTML, JSON, XML.

Tabla 1
Parámetros en la base de datos

ID
Nombre de la red social
Nombre del lugar
Comentario más nuevo
Puntaje positivo global
Puntaje negativo global
Número de comentarios positivos total
Número de comentarios negativos
Puntaje positivo en base a un parámetro
Puntaje negativo en base a un parámetro
Número de comentarios positivos en base a un parámetro
Número de comentarios negativos en base a un parámetro

Nota: Campos que permitieron almacenar información de redes sociales

Fase de Análisis

Con la información proveniente de las dos fuentes se realizó el análisis de sentimientos utilizando la librería *Sentiment* que tiene una lista aproximada de 3382 palabras en inglés. Se tradujo cada una de las palabras con Google Translate usando *Puppeteer* y *Node.js*. A esta librería se incluyeron 4 parámetros: precio, servicio, transporte y hospedaje. Con la finalidad de ampliar el diccionario que permita clasificar y analizar los comentarios emitidos, hubo la necesidad de expandir las palabras para lo cual se utilizó el sitio web (*Synonimos*, 2022), posteriormente se eliminaron las palabras repetidas generadas en este proceso.

Los parámetros añadidos en este proceso incluyeron sinónimos como muestra la Figura 2. Que permita extender la búsqueda de comentarios que tengan una de estas palabras y obtener un puntaje más robusto a base del precio, servicio, transporte y hospedaje.

Figura 2
Uso de la Librería Sentiment

```
{
  score: -3,
  comparative: -0.15,
  tokens:
    [ 'el',
      'tweet',
      'que',
      'aparece',
      'abajo',
      'es',
      'falso',
      'el',
      'alcalde',
      '@mauriciorodasec',
      'ha',
      'reiterado',
      'su',
      'total',
      'independencia',
      'frente',
      'al',
      'actual',
      'proceso',
      'electoral' ],
  words: [ 'proceso', 'actual', 'frente', 'total', 'falso', 'abajo' ],
  positive: [ 'actual', 'frente', 'total' ],
  negative: [ 'proceso', 'falso', 'abajo' ] }
```

Nota: Código Node js

Finalmente, al realizar el análisis de sentimientos en las palabras existentes en el comentario, se obtuvo un puntaje. Si el comentario tiene un puntaje mayor a 0 es positivo, si es menor a 0 es negativo y si es igual a 0 es neutro.

Fase de Visualización

Es necesario presentar gráficos que se actualicen en tiempo real. El sistema desarrollado utilizó como herramienta de visualización JavaScript y Chart.js debido a que las dos se pueden integrar y presentar los datos de mejor manera. Además, se utilizó la librería Socket.io que permite actualizar gráficos en tiempo real y poder tomar decisiones de mejoras o creación de nuevos servicios.

Los gráficos elaborados contemplan los siguientes resultados:

1. Porcentajes de comentarios positivos, negativos y neutros de servicios ofertados por lugares turísticos.
2. Resultados estadísticos agrupados por las redes sociales: Twitter y Trip Advisor.
3. Cuadro resumen de porcentajes de los sentimientos obtenidos en los comentarios emitidos por los usuarios.

Selección de las herramientas

Para la elección de las herramientas adecuadas se ha utilizado el método de criterios ponderados. Este método se basa en comparar las herramientas mediante la generación de criterios de ingeniería que son propuestos según los requerimientos del sistema. Se busca comparar todas entre sí para determinar el impacto de cada una frente a otra, de manera que se conoce qué variables tienen mayor impacto y cuáles deben ser analizadas con cuidado en el área de diseño (Kalpakjian & Schmid, 2016).

La etapa de definición de los criterios comenzó por identificar los parámetros que se consideran fundamentales para el desarrollo del sistema, basado en los requerimientos y la experiencia en el uso de diferentes herramientas informáticas. En función a este método se detalla el análisis y selección de herramientas utilizadas en las fases del Big Data.

Recolección de datos

Las herramientas disponibles orientadas a la recolección de datos y que fueron evaluadas: twitterR (R-Studio), Puppeteer (Node.js) y Twython (Python). Estas librerías se consideran en función de la independencia de una API que estas tengan, ya que es necesario extraer una gran cantidad de comentarios sin perder información.

Para la selección de la herramienta se consideró los siguientes criterios basándose en librerías de diferentes lenguajes de programación desde un punto de vista de ingeniería: independencia de una API y extracción completa de comentarios, la Tabla 2 muestra el análisis de las diferentes opciones.

Según las calificaciones obtenidas la mejor herramienta para la recolección de datos fue: Puppeteer. Esto se debe a que la librería permite extraer comentarios navegando a la red social por medio de la programación y el uso del navegador Chromium (Google Developers, 2021). Además, esta librería es totalmente independiente de una API y permite obtener comentarios completos por medio del código HTML de la red social.

Tabla 2*Matriz de Criterios Ponderados: Fase de Recolección de Datos*

	<i>LIBRERÍAS PARA RECOLECTAR COMENTARIOS</i>	<i>PONDERACIÓN</i>	<i>TWITTER</i>	<i>PUPPETEER</i>	<i>TWYTHON</i>
	<i>CRITERIO</i>				
Relevantes	<i>Independencia de una API</i>	<i>60</i>	<i>40</i>	<i>55</i>	<i>48</i>
Efectivos	<i>Extracción Completa de Comentarios</i>	<i>40</i>	<i>25</i>	<i>37</i>	<i>30</i>
<i>Calificación</i>		<i>100</i>	<i>65</i>	<i>92</i>	<i>78</i>

Nota: Análisis de herramientas para la recolección de datos

Procesamiento de Datos

Las herramientas de procesamiento que fueron analizadas: Sentiment (R-Studio), Sentiment (Node.js) y Vader (Python). Estas herramientas se seleccionaron a base del análisis de sentimientos que se va a realizar de los comentarios que provienen de redes sociales.

Para el procesamiento de datos se plantearon los siguientes criterios, basados en librerías de diferentes lenguajes de programación desde un punto de vista de ingeniería: diccionario extenso de palabras, sistema de evaluación de la librería y documentación disponible.

Tabla 3*Matriz de Criterios Ponderados: Fase de Procesamiento de Datos*

	<i>LIBRERÍAS PARA EL ANÁLISIS DE SENTIMIENTOS</i>	<i>PONDERACIÓN</i>	<i>SENTIMENT (R-STUDIO)</i>	<i>SENTIMENT (NODE.JS)</i>	<i>VADER (PYTHON)</i>
	<i>CRITERIO</i>				
Relevantes	<i>Diccionario Extenso de Palabras</i>	<i>50</i>	<i>35</i>	<i>40</i>	<i>38</i>
Efectivos	<i>Sistema de Evaluación de la Librería</i>	<i>35</i>	<i>20</i>	<i>32</i>	<i>28</i>
Flexibles	<i>Documentación Disponible</i>	<i>15</i>	<i>8</i>	<i>14</i>	<i>12</i>
<i>Calificación</i>		<i>100</i>	<i>63</i>	<i>86</i>	<i>78</i>

Nota: Análisis de herramientas para el procesamiento de datos

De acuerdo a los resultados obtenidos la mejor herramienta para el procesamiento de datos fue la librería Sentiment del lenguaje de programación Node.js. Su puntaje se debe a que

fue necesario crear un código para expandir el diccionario de la librería (Node Js, 2022). Este código parte del concepto de añadir sinónimos de las palabras que existen por defecto usando el sitio web sinonimos.com.

Análisis y almacenamiento de datos (Si el análisis se aplicó en el estudio, se redacta en tiempo pasado)

Las herramientas que se evaluaron para el análisis de datos: R-Studio, Node.js y Python. Los criterios se tomaron en base a: las librerías disponibles como un criterio relevante, la curva de aprendizaje del lenguaje como un criterio flexible ya que es necesario tener una buena cantidad de librerías para poder crear el sistema de manera eficaz, integración con el lenguaje de programación y finalmente consistencia.

Tabla 4

Matriz de Criterios Ponderados: Fase de Análisis de Datos

	<i>LENGUAJE DE PROGRAMACIÓN (BACK-END)</i>	<i>PONDERACIÓN</i>	<i>R-STUDIO</i>	<i>NODE.JS</i>	<i>PYTHON</i>
	<i>CRITERIO</i>				
<i>Relevantes</i>	<i>Librerías Disponibles</i>	<i>70</i>	<i>55</i>	<i>68</i>	<i>60</i>
<i>Flexibles</i>	<i>Curva de Aprendizaje</i>	<i>35</i>	<i>30</i>	<i>30</i>	<i>22</i>
	<i>Calificación</i>	<i>100</i>	<i>81</i>	<i>91</i>	<i>82</i>

Nota: Análisis de herramientas para el análisis de datos

De acuerdo a los resultados obtenidos la mejor herramienta para el análisis de datos fue: Node.js. Esto debido a que el lenguaje tiene una cantidad considerable de librerías. Algunas de estas librerías incluyen a Puppeteer y Socket.io y librerías que permiten el manejo de la base de datos SQL o NoSQL.

En la Tabla 5 se muestran los resultados obtenidos después de evaluar cada uno de los criterios considerados en la selección de la base de datos NoSQL.

Según las calificaciones obtenidas la mejor herramienta para el almacenamiento de datos fue base de datos NoSQL MongoDB. Esto se debe a que MongoDB tiene una fácil integración con el lenguaje Node.js y existen varias librerías que ayudan a que MongoDB tenga más funcionalidad. Además, la aplicación va a ser en tiempo real por lo que se necesita reflejar los datos inmediatamente.

Tabla 5*Matriz de Criterios Ponderados: Fase de Almacenamiento de Datos*

	<i>BASE DE DATOS NOSQL</i>	<i>PONDERACIÓN</i>	<i>CASSANDRA</i>	<i>MONGODB</i>
	<i>CRITERIO</i>			
<i>Relevantes</i>	<i>Integración con el Lenguaje de Programación</i>	<i>65</i>	<i>50</i>	<i>60</i>
<i>Flexibles</i>	<i>Consistencia</i>	<i>35</i>	<i>30</i>	<i>30</i>
<i>Calificación</i>		<i>100</i>	<i>80</i>	<i>90</i>

Nota: Análisis de herramientas para el almacenamiento de datos

Visualización de datos

Las herramientas que fueron analizadas para poder visualizar datos: R- Studio, Python, JavaScript y Chart.js. Los criterios que se tomaron en consideración para la fase de visualización de datos se basaron en la utilidad para visualizar datos como un criterio relevante y la facilidad de uso como un criterio flexible ya que es necesario tener una interfaz que permita llegar a conclusiones sobre el destino turístico a través de los gráficos. Los criterios que determinaron la herramienta para la visualización se describen: facilidad de visualizar datos y curva de aprendizaje.

Tabla 6*Matriz de Criterios Ponderados: Fase de Visualización de Datos*

	<i>LENGUAJE DE PROGRAMACIÓN (FRONT-END)</i>	<i>PONDERACIÓN</i>	<i>R-STUDIO</i>	<i>PYTHON</i>	<i>JAVASCRIPT</i>	<i>CHART.JS</i>
	<i>CRITERIO</i>					
<i>Relevantes</i>	<i>Facilidad de Visualizar Datos</i>	<i>65</i>	<i>55</i>	<i>55</i>	<i>57</i>	<i>63</i>
<i>Flexibles</i>	<i>Curva de Aprendizaje</i>	<i>35</i>	<i>20</i>	<i>33</i>	<i>34</i>	<i>33</i>
<i>Calificación</i>		<i>100</i>	<i>75</i>	<i>88</i>	<i>91</i>	<i>96</i>

Nota: Análisis de herramientas para la visualización de datos

Las mejores herramientas para la visualización de datos fueron JavaScript y Chart.js debido a que las dos se pueden integrar y presentar los datos de una manera elegante. Además, se puede usar con la librería Socket.io (Char Js, 2022).

Redes Sociales

Al momento de crear un sistema que evalúe y recomiende destinos turísticos es importante considerar que redes sociales son adecuadas para proveer datos. Algunas redes sociales tienen mejor calidad de datos, otras tienen más seguridad y otras requieren que el usuario se registre para poder visualizar comentarios por lo que es de gran importancia tomar esto en cuenta.

Las redes sociales que se compararon: Facebook, Twitter, Trip Advisor e Instagram. Estas redes sociales fueron seleccionadas por ser las más utilizadas y relevantes. Se incluyó a Trip Advisor pues al momento es considerada una red social exclusiva para obtener comentarios de lugares turísticos y la información podría resultar valiosa para realizar el estudio establecido.

Para la selección de redes sociales se plantearon los siguientes criterios desde un punto de vista de ingeniería son: alta calidad de datos, baja protección de datos, bajo nivel de seguridad, cantidad de comentarios diarios y número de usuarios.

Tabla 7

Matriz de Criterios Ponderados: Redes Sociales

	<i>RED SOCIAL</i>	<i>PONDERACIÓN</i>	<i>FACEBOOK</i>	<i>TWITTER</i>	<i>INSTAGRAM</i>	<i>TRIP ADVISOR</i>
	<i>CRITERIO</i>					
<i>Relevantes</i>	<i>Alta Calidad de Datos</i>	<i>80</i>	<i>22</i>	<i>25</i>	<i>25</i>	<i>28</i>
	<i>Bajo Nivel de Seguridad</i>		<i>18</i>	<i>22</i>	<i>20</i>	<i>23</i>
	<i>Bajo Nivel de Seguridad</i>		<i>15</i>	<i>13</i>	<i>8</i>	<i>11</i>
<i>Efectivos</i>	<i>Cantidad de comentarios diarios</i>	<i>15</i>	<i>15</i>	<i>13</i>	<i>34</i>	<i>33</i>
<i>Flexible</i>	<i>Número de usuarios</i>	<i>5</i>	<i>5</i>	<i>5</i>	<i>3</i>	<i>3</i>
<i>Calificación</i>		<i>100</i>	<i>75</i>	<i>87</i>	<i>76</i>	<i>89</i>

Nota: Análisis de redes sociales para recolección de datos

Las redes sociales más aptas para obtener grandes volúmenes de comentarios fueron Twitter y Trip Advisor. Trip Advisor es la herramienta que cuenta con la mejor calificación en la calidad de datos porque esta red social tiene comentarios que se relacionan a lugares turísticos. Twitter en cambio tiene un mayor número de usuarios y comentarios que serán de gran ayuda al momento de evaluar un destino turístico. Sin embargo, los comentarios de la red social Twitter provienen de las cuentas que pertenecen a un solo usuario y por este motivo existen comentarios que posiblemente sean neutros.

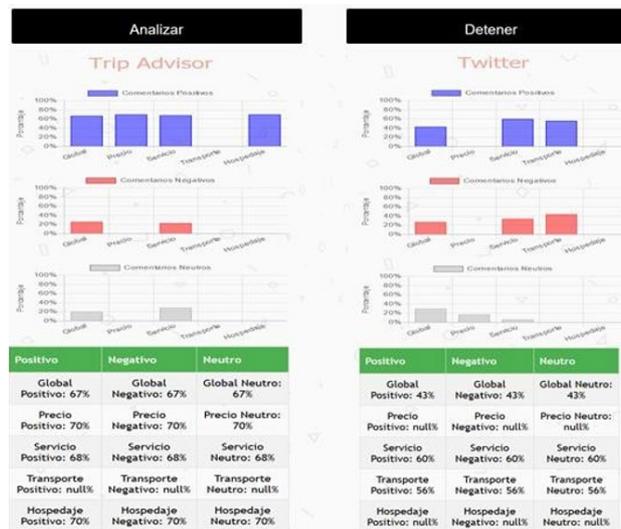
Resultados y Discusión

Como resultado del análisis realizado, la Figura 3, muestra el resultado del análisis de sentimientos que corresponde al sitio turístico Mitad del Mundo; basado en la utilización de la

librería Sentiment, el sitio a ser evaluado, y los parámetros establecidos como: el precio, servicio, transporte y hospedaje.

Figura 3

Análisis de sentimientos Twitter y Trip Advisor



Nota. Información resultante del proceso de análisis de sentimientos

Como se puede observar los comentarios difieren de las fuentes de datos utilizados en el análisis. En el grupo de comentarios positivos se determina que mientras el sitio Trip Advisor no muestra registros en el parámetro transporte, la red social Twitter si contiene registros. En el grupo de comentarios negativos se puede observar que las dos fuentes de datos muestran información penalizando los parámetros: servicio y transporte. Finalmente, se puede observar que existen comentarios neutros asociado con los parámetros: precio y servicio.

El sistema presenta un cuadro de resumen que muestra el total de comentarios positivos, negativos y neutros, emitidos en las dos fuentes de datos. Uno de los puntos importantes de la aplicación es poder presentar en tiempo real los comentarios emitidos, para lo cual es necesario capturar datos, almacenarlos y presentar una gráfica con comentarios positivos, negativos y neutros, esta información permite determinar posibles puntos de mejora en los servicios, tanto con la información del sitio Trip Advisor, como de los tuits, mediante la interfaz gráfica de la aplicación.

De la información conseguida, actualmente este sitio dispone de encuestas como herramienta para evaluar la calidad de los servicios ofertados. La mayoría no disponen de un sistema que permita observar en tiempo real información de sentimientos asociados a la conformidad en la utilización de los productos que el lugar oferta.

Finalmente, de los resultados obtenidos en la presente investigación permite argumentar:

El estudio realizado tiene una limitante considerable relacionada con la necesidad de la activación de la opción de geolocalización en las cuentas de Twitter a ser utilizadas.

Es importante disponer de un sistema en tiempo real que permita determinar puntos de mejora en los servicios que son ofrecidos por los lugares turísticos, para de esta manera tomar decisiones adecuadas y basadas en información real.

Al ser Big Data una de las nuevas tecnologías de la revolución industrial 4.0, es utilizada y aplicada en la resolución de problemas de cualquier ámbito del conocimiento en el ámbito mundial, el turismo no es la excepción por lo cual es apropiado el uso de esta tecnología en el contexto de la presente investigación.

Conclusiones

La investigación realizada determina la necesidad de contar con herramientas que permitan tomar decisiones, para el mejoramiento de la calidad de los servicios ofertados por las empresas del sector turístico. Por otro lado, se debe mencionar que la utilización de datos generados en Twitter y Trip Advisor, han permitido la creación de un sistema que permita analizar los sentimientos emitidos por usuarios de servicios turísticos. Se puede, además, considerar la utilización del sistema desarrollado, como una alternativa viable y de menor costo que permita la evaluación y mejora de servicios relacionados con el turismo. De acuerdo al contexto de la investigación el uso de una base de datos NoSQL en Big Data es importante para resolver la problemática de escalabilidad, que generalmente se presenta en bases de datos relacionales, debido a la gran cantidad de información que se necesite almacenar.

La red social Trip Advisor tiene mejor calidad de datos que Twitter con base en destinos turísticos, ya que los comentarios provienen de la opinión de varios usuarios. De la misma manera, la red social Twitter es de mayor utilidad al momento de realizar el análisis de sentimientos a base de los temas sociales, motivo por el cual el resultado en esta red social tiene una mayor cantidad de comentarios neutrales con relación a destinos turísticos.

Tanto Twitter como Trip Advisor tienen una gran cantidad de bots y cuentas falsas, por lo que sería importante realizar un estudio que permita establecer un filtro y excluir esa información de investigaciones, que toman como fuente de información las redes sociales mencionadas. Finalmente, se precisa que una de las grandes ventajas de haber realizado este sistema es que al momento de planificar se podría saber si una base de datos relacional resulta más conveniente usar que una NoSQL y viceversa.

Reconocimientos

Esta investigación se enmarca en el proyecto de investigación “Evaluación y promoción de destinos turísticos, mediante la tecnología de la ciencia de datos y geolocalización” apoyado por el Vicerrectorado del Instituto Superior Tecnológico Japón.

Referencias

- Aguilar, L. J. (2013). *Análisis de grandes volúmenes de datos en organizaciones*. Mexico: Alfaomega Grupo Editor, S.A.
- Algecira, C. (2020). *Arquitectura de Big Data para el análisis de sentimientos multimodales en el sector turístico del Departamento de Boyacá, Colombia*. Boyacá: Universidad Santo Tomás Sede Tunja.
- Ángel Patricio Fajardo Cárdenas, Néstor Ariel Bravo Chuqui, Andrés Vinicio Auquilla Sangolqui, & Paúl Fernando Vanegas P. (2021). *Plataforma para el Análisis de Mercado a través de Datos de Redes Sociales*. Guayaquil: Resvista RTE.
- Arias, Á. (2017). *El uso de la plataforma trip advisor. Estudio de caso: Los comentarios sobre los Hoteles en Valladolid*. Valladolid: Universidad de Valladolid.
- Char Js. (01 de 02 de 2022). *Performance*. Performance: <https://www.chartjs.org/docs/latest/general/performance.html>

- Escribano, Á. A. (2017). *El uso de la herramienta Trip Advisor. Estudio de Caso: Los comentarios sobre hoteles de Valladolid*. Segovia: Universidad de Valladolid.
- Fajardo, Á., Auquilla, A., & Vanegas, P. (2021). *Plataforma para Análisis de Mercado a través de Datos de Redes Sociales*. Guayaquil: Revista Tecnológica Espol.
- Google Developers. (03 de 02 de 2021). *Tools for Web Developers*. Tools for Web Developers: <https://developers.google.com/web/tools/puppeteer/>
- Herrera, N., Sánchez, H., Macías, M., & Gómez, E. (2016). *Analysis of the factors generating vehicular traffic in the city of Quito and its relation to the application of sensorial and social data with big data as a basis for decision making*.
- Kalpakjian, S., & Schmid, S. (2016). *Manufactura, ingeniería y tecnología*. Mexico: Pearson.
- Krishna, S. (01 de 07 de 2018). *Facebook puts more limits on developer access to user data*. Facebook puts more limits on developer access to user data: https://www.engadget.com/2018-07-02-facebook-limit-api-developer-app-data.html?sr_source=Twitter
- Letelier, P., & Penadés, M. (2016). (XP), Metodologías ágiles para el desarrollo de software: eXtreme Programming. *Técnica Administrativa*.
- Martínez, X., & Piñeiro, T. (2020). *Turismo y Big Data. Propuesta de Análisis de la capa social a través de metadatos y publicaciones sociales*. Portugal: Cultura, patrimonio e turismo na sociedade de digital.
- Ministerio de Turismo del Ecuador. (2019). *Plan Estratégico Institucional 2019-2021*. Ministerio de Turismo del Ecuador.
- Node Js. (15 de 01 de 2022). *Dependencias*. Dependencias: <https://nodejs.org/es/docs/>
- Padilla, C. (2019). *Big Data, una herramienta para apoyar en decisiones del sector hotelero en Quito-Ecuador*. Quito: Revista UIDE.
- Rik, J., & Tim de Boer. (2011). *Twitter An Architectural Review*. Utrecht: Software Architecture Faculty of Science University of Utrecht. Netherland: Faculty of Science University of Utrecht.
- Synonimos. (12 de 02 de 2022). *Synonimos*. Synonimos: <http://www.sinonimos.com/>
- Tello, J. M. (2019). *Potencial aplicación del Big Data al sector turismo con especial referencia a España*. Madrid: Universidad Pontificia Comillas.
- Twitter. (2022). *Centro de ayuda*. Información sobre las API de Twitter: <https://help.twitter.com/es/rules-and-policies/twitter-api>
- Zicopoulos, P., Chris, E., Dirk, D., Tom, D., & George, L. (2011). *Analytics for Enterprise Class Hadoop and Streaming Data*. McGraw-Hill Osborne Media.