

## Análisis de Clusterización en Datos de Encuestas sobre Ciberacoso

### Clustering Analysis on Cyberbullying Survey Data

César Vinueza-Álvarez<sup>1</sup> <https://orcid.org/0009-0001-6731-3314>, María Inés Acosta-Urigüen<sup>1</sup> <https://orcid.org/0000-0003-4865-2983>, Juan Fernando Lima<sup>1</sup> <https://orcid.org/0000-0003-3500-3968>

<sup>1</sup>*Departamento de Posgrados, Universidad del Azuay, Cuenca, Ecuador*  
[cvinuezamsn@es.uazuay.edu.ec](mailto:cvinuezamsn@es.uazuay.edu.ec)

<sup>2</sup>*Laboratorio de Investigación y Desarrollo en Informática (LIDI), Universidad del Azuay, Cuenca, Ecuador*  
[macosta@uazuay.edu.ec](mailto:macosta@uazuay.edu.ec), [flima@uazuay.edu.ec](mailto: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.

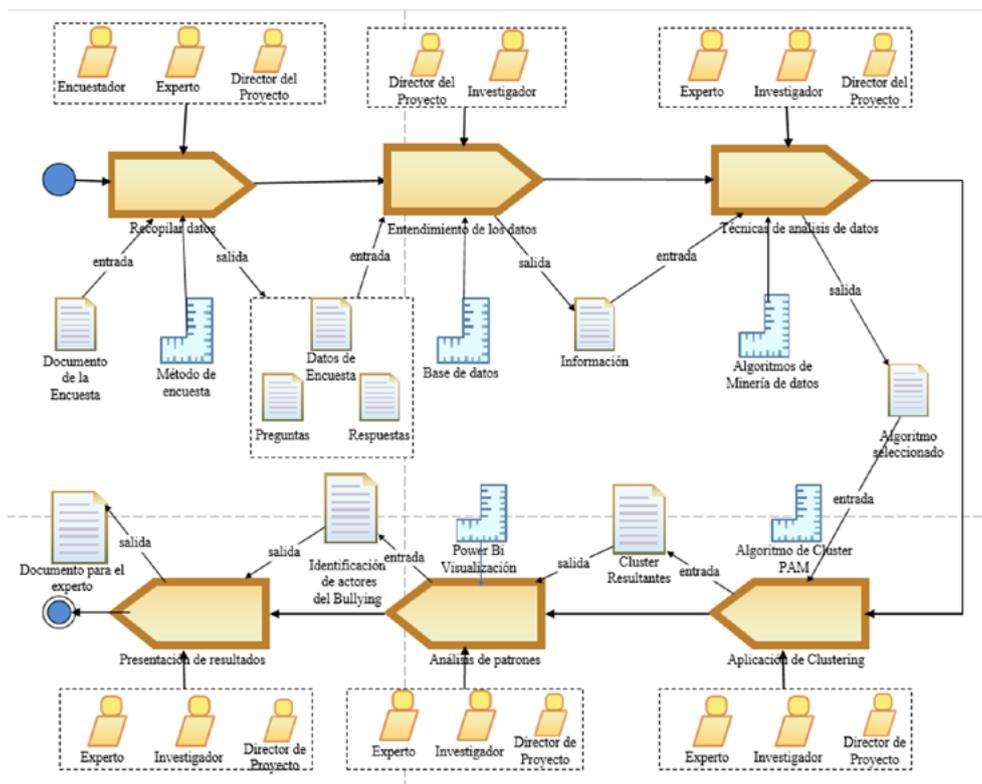
Los autores Dedic et al. (2019) analizan los datos obtenidos con tres tipos de encuestas estandarizadas (“*The Juvenile Victimization Questionnaire*”, “*The Defense Style Questionnaire*” y “*Beck Depression Inventory*”) realizadas a estudiantes del Sur de Australia, Gran Bretaña y Suecia a través de 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.

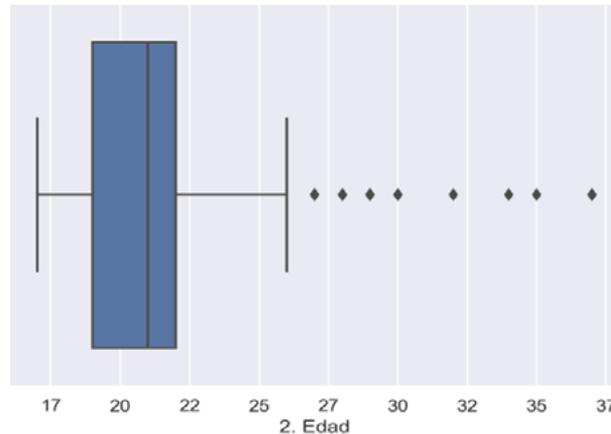
Los datos recolectados en la encuesta corresponden a un grupo de adolescentes con una edad entre los 18 y 22 años, con un nivel socioeconómico medio alto, que asisten a una universidad de la ciudad de Cuenca, Ecuador. Se aplicó la encuesta considerando una población de 7,000 estudiantes con un nivel 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
“Cielo”	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

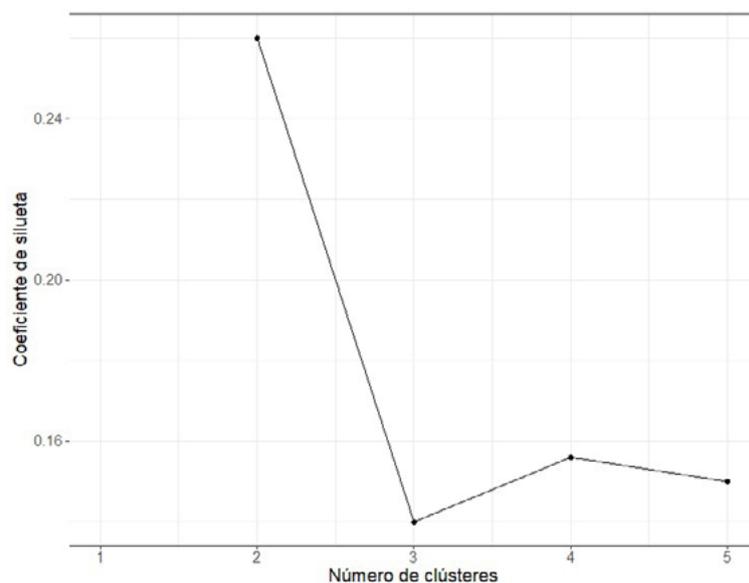
### Definición y aplicación del modelo

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

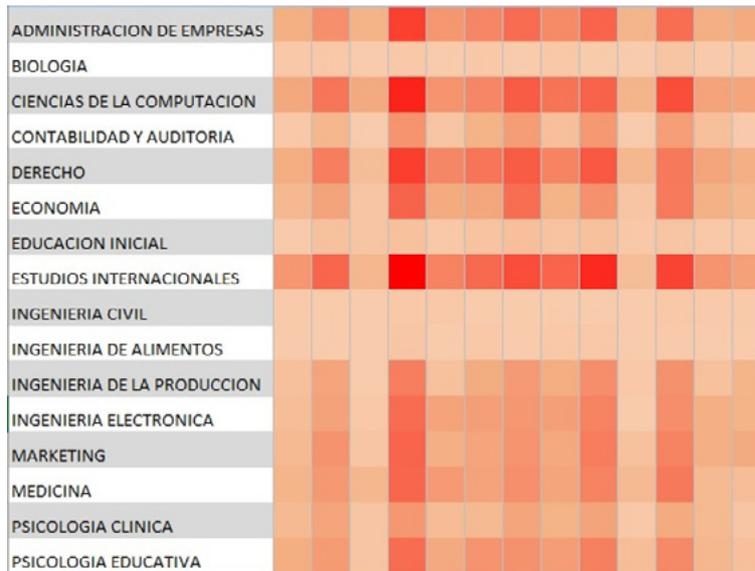


Una vez que se estableció el número de clústeres, se aplicó el algoritmo de asignación del clúster para cada uno de los individuos. Se empleó la distancia de Gower, para encontrar las medidas entre los sujetos estudiados; y se utilizó el algoritmo PAM, para la generación de los clústeres. Este algoritmo 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

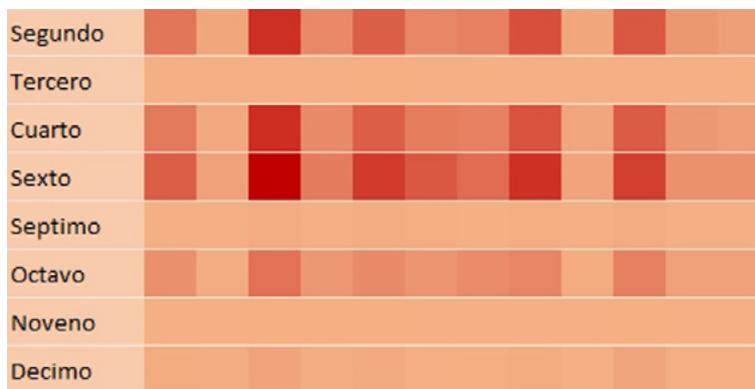
En esta parte se validó el análisis de los clústeres realizados para luego evaluar los resultados obtenidos. Para lo cual, se elaboraron varios modelos que poseen la suficiente información desde una perspectiva de análisis de datos, de tal modo que se puedan analizar y verificar los resultados de manera 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.

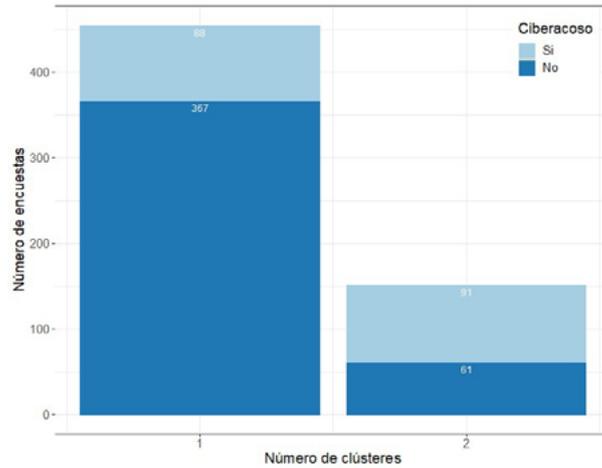
**Figura 5**  
*Nivel de acoso y ciberacoso por ciclo académico*



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.

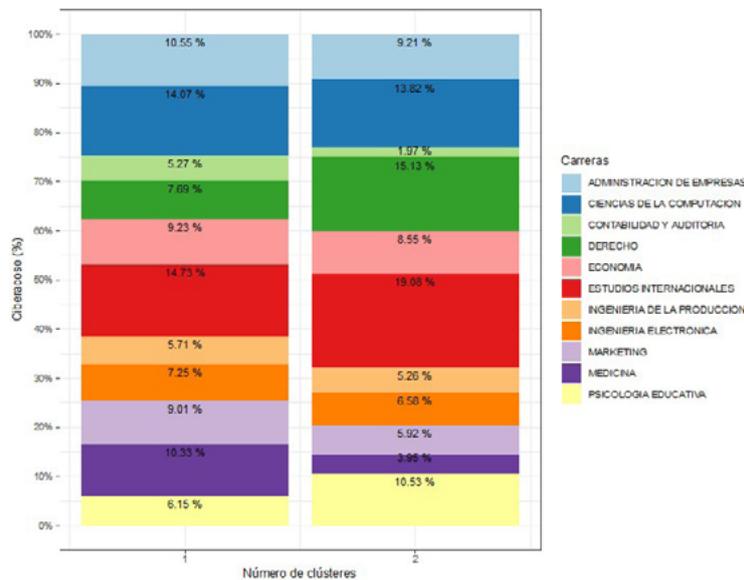
**Figura 6**  
Cantidad de personas que recibieron ciberacoso



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.

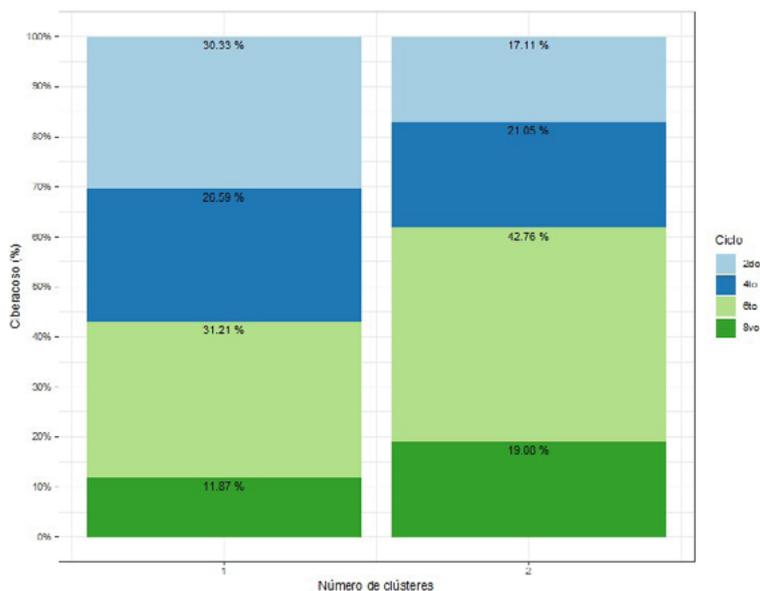
Una de las primeras variables que se analizó fue las carreras de la universidad que presentaron un mayor índice de ciberacoso en base a las características de la agrupación de los clústeres. Por lo tanto, existe una mayoría relativa en ambos clústeres con las carreras de Estudios Internacionales, Derecho 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>