







## Minería de datos y redes sociales: predicción de la distracción en estudiantes de educación superior

### Data Mining and Social Media: Predicting Distraction Among College Students

Alan Ramírez-Noriega <sup>1</sup>, Yobani Martínez-Ramírez <sup>2</sup>, Juan Francisco Figueroa Pérez <sup>3</sup>,  
Carolina Tripp-Barba <sup>4</sup> y Marcos Adrián Murillo-Corrales <sup>5</sup>

**Fecha de Recepción:** 12 de noviembre de 2025

**Fecha de Aceptación:** 28 de febrero de 2026

**Cómo citar:** A. Ramírez-Noriega, Y. Martínez-Ramírez, J. F. Figueroa Pérez, C. Tripp-Barba, y M. A. Murillo-Corrales, «Minería de datos y redes sociales: predicción de la distracción en estudiantes de educación superior» *Tecnura*, vol. 30, n.º 87, mar. 2026. 1–18. <https://doi.org/10.14483/22487638.24842>

## Resumen


**Objetivo:** las redes sociales, como herramientas emergentes de las últimas tres décadas, han evolucionado en paralelo con Internet. Estas plataformas ofrecen formas de interacción que pueden derivar en distracciones positivas como negativas, según el uso que se les dé. En el ámbito educativo, un uso inadecuado de estas plataformas digitales puede generar problemas de distracción y afectar negativamente el aprendizaje. En este contexto, la presente investigación desarrolla un modelo predictivo basado en minería de datos y aprendizaje automático, con el objetivo de identificar el grado de distracción en el proceso de enseñanza-aprendizaje.


**Metodología:** la base de conocimiento, construida a partir de una encuesta aplicada a más de 1000 estudiantes de educación superior, proporciona la información necesaria para entrenar algoritmos de aprendizaje automático capaces de predecir la probabilidad de que un estudiante sufra distracciones debido al uso de las redes sociales.


**Resultados:** los resultados obtenidos muestran que el modelo es capaz de detectar adecuadamente el 80 % de los casos de distracción, aunque existe margen para mejorar estos resultados mediante ajustes en la base de conocimiento.


**Conclusiones:** los hallazgos actuales demuestran un buen rendimiento en la predicción de distracciones relacionadas con el uso de las plataformas digitales, lo que permitiría anticipar el bajo rendimiento de estudiantes y proponer estrategias de mejora.


**Palabras clave:** Minería de datos; Redes sociales; Predicción; Estudiante.

1 Profesor e Investigador de Tiempo Completo. Facultad de Ingeniería Mochis, Universidad Autónoma de Sinaloa. Email: [alandramireznoriega@uas.edu.mx](mailto:alandramireznoriega@uas.edu.mx)

2 Profesor e Investigador de Tiempo Completo. Facultad de Ingeniería Mochis, Universidad Autónoma de Sinaloa. Email: [yobani@uas.edu.mx](mailto:yobani@uas.edu.mx)

3 Profesor e Investigador de Tiempo Completo. Facultad de Ingeniería Mochis, Universidad Autónoma de Sinaloa. Email: [juanfc.figueroa@uas.edu.mx](mailto:juanfc.figueroa@uas.edu.mx)

4 Profesora e Investigadora de Tiempo Completo. Facultad de Informática Mazatlán, Universidad Autónoma de Sinaloa. Email: [ctripp@uas.edu.mx](mailto:ctripp@uas.edu.mx)

5 Profesor e Investigador de Tiempo Completo. Facultad de Agricultura del Valle del Fuerte, Universidad Autónoma de Sinaloa. Email: [avancemarcos@uas.edu.mx](mailto:avancemarcos@uas.edu.mx)

## Abstract

**Objective:** Social media platforms, as emerging tools over the last three decades, have evolved alongside the Internet. They offer forms of interaction that can lead to positive or negative distractions, depending on how they are used. In the educational context, the inappropriate use of these digital platforms can lead to distraction issues and negatively affect learning. In this context, this research develops a predictive model based on data mining and machine learning, aiming to identify the degree of distraction in the teaching-learning process.

**Methodology:** The knowledge base, constructed from a survey administered to over 1,000 higher education students, provides the necessary information to train machine learning algorithms capable of predicting the probability of a student experiencing distraction due to social media use.

**Results:** The results show that the model accurately detects 80% of distraction cases, although there is room for improvement through adjustments to the knowledge base.

**Conclusions:** The findings demonstrate good performance in predicting distractions related to the use of digital platforms, which would allow anticipating low academic performance in students and proposing intervention strategies.

**Keywords:** Data mining; Social media; Prediction; Student

---

## Introducción

Las redes sociales revolucionaron la forma en que nos comunicamos e interactuamos y transformaron profundamente diversos aspectos de nuestra vida, incluida la educación. Estas plataformas digitales, que conectan a millones de personas en todo el mundo, ofrecen una amplia gama de herramientas y recursos que pueden ser utilizados de manera efectiva en el ámbito académico. Sin embargo, su impacto en el aprendizaje es complejo y multifacético, con evidencias tanto de beneficios como de desafíos en su aplicación [1].

Por un lado, las redes sociales pueden ser una herramienta poderosa para el aprendizaje colaborativo, el acceso a información y el desarrollo de habilidades digitales. Estas plataformas permiten a los estudiantes conectar con compañeros, profesores y expertos en todo el mundo, lo que facilita el intercambio de ideas, la resolución de problemas y la adquisición de conocimientos de manera más dinámica y atractiva [2]. Además, pueden servir como un complemento valioso a los métodos de enseñanza tradicionales, al ofrecer nuevas oportunidades para la personalización del aprendizaje y la creación de comunidades de aprendizaje en línea [3].

No obstante, el uso excesivo o inadecuado de las redes sociales también puede tener consecuencias negativas para el rendimiento académico [4]. La constante exposición a estímulos y distracciones en línea puede dificultar la concentración y la atención, lo que a su vez puede afectar la capacidad de los estudiantes para retener información y completar tareas. Asimismo, el uso de las redes sociales puede generar problemas de aislamiento social, ansiedad y depresión, especialmente cuando se utiliza como un sustituto de las interacciones sociales cara a cara [5].

La proliferación de redes sociales ha dado lugar a una amplia gama de distracciones que pueden desviar la atención de los estudiantes de sus responsabilidades académicas. Las notificaciones constantes, los mensajes y las actualizaciones en las comunidades virtuales pueden interrumpir el flujo de trabajo y generar una sensación de urgencia por responder de inmediato. Además, el fácil acceso a contenido multimedia, como videos y juegos, puede resultar tentador para los estudiantes, lo que a menudo conduce a una pérdida de tiempo y una disminución de la productividad [6], [7].

Los estudiantes que se distraen con frecuencia con las redes sociales pueden enfrentar una serie de problemas académicos y personales: dificultades para cumplir con los plazos de entrega, bajo rendimiento en exámenes, problemas de gestión del tiempo, y deterioro de las relaciones interpersonales. A largo plazo, estos problemas pueden afectar negativamente su desarrollo académico y profesional [8].

El objetivo de este proyecto es determinar el grado de distracción de los estudiantes con mediante un instrumento de encuesta. Se busca predecir el grado de distracción con el fin de proponer alternativas previas a los problemas derivados del uso de redes sociales. Para ello se emplean algoritmos predictivos del aprendizaje máquina y un instrumento aplicado a más de mil estudiantes con el propósito de construir la base de conocimiento para la toma de decisiones.

El artículo está estructurado de la siguiente manera. La segunda sección describe el marco teórico con los conceptos relacionados más importantes de la investigación. La tercera sección expone la metodología empleada con el instrumento para recolectar información y los procesos de minería empleados; además se muestran los resultados obtenidos. Posteriormente se muestra la discusión de los resultados y las conclusiones. Finalmente, aparecen las referencias que brindan soporte a la investigación.

## Marco Teórico

A continuación, se muestran los conceptos principales de la investigación.

### 2.1 KDD

El Descubrimiento del Conocimiento en Base de Datos (*Knowledge Discovery Databases*, KDD por sus siglas en inglés) es el proceso no trivial de identificar patrones válidos, novedosos, potencialmente útiles y comprensibles a través de los datos [9]. La frase descubrimiento de conocimiento en bases de datos se acuñó en el primer taller de KDD en 1989 por Piatetsky-Shapiro para enfatizar que el conocimiento es el producto final de un descubrimiento impulsado por datos. Se ha popularizado en los campos de la inteligencia artificial y el aprendizaje automático [10].

El proceso de KDD se organiza en cinco fases [11], [12]:

- a. Integración y recopilación: se determinan las fuentes de información que pueden ser útiles y dónde conseguirlas. A continuación, se transforman los datos a un formato común mediante un almacén de datos que consiga unificar de manera operativa toda la información recogida, al tiempo que se detectan y corrigen las inconsistencias.
- b. Selección, limpieza y transformación: dado que los datos provienen de diferentes fuentes, pueden contener valores erróneos o faltantes. Estas situaciones se tratan en esta fase, en la cual se eliminan o corrigen los datos incorrectos y se decide la estrategia a seguir con los datos incompletos.
  1. Selección: se proyectan los datos para considerar únicamente aquellas variables o atributos que van a ser relevantes, con el objetivo de hacer más fácil la tarea de minería. Si son muchos datos, se debe considerar el uso del muestreo, seleccionando solo algunas de las filas, con el objetivo de agilizar el procesamiento de información.
  2. Limpieza: se detecta la presencia de datos faltantes o perdidos (*missing values*) que pueden conducir a resultados poco precisos. También se detectan valores que no se ajustan al comportamiento general de los datos (*outliers*).
  3. Transformación: consiste en construir automáticamente nuevos atributos aplicando alguna operación o función a los atributos originales con el objeto de que estos atributos hagan más fácil el proceso de minado.
- c. Minería de datos: el objetivo de esta etapa es la búsqueda y descubrimiento de patrones insospechados y de interés. Se toman varias decisiones como: 1) determinar qué tipo de tarea es la más apropiada (clasificación, regresión, clusterización, etc.). 2) elegir el tipo de modelo (redes neuronales, arboles de decisión, Bayes ingenuo, entre otros). 3) elegir el algoritmo de minería que resuelva la tarea y obtenga el tipo de modelo que estamos buscando.
- d. Evaluación e interpretación: se evalúan e interpretan los patrones descubiertos y, posiblemente, se retorna a las anteriores etapas para posteriores iteraciones. La evaluación depende del tipo de tarea aplicada.
- e. Difusión, uso y monitorización: la última fase es la implementación del sistema desarrollado, así como la inclusión de una fase de mantenimiento que permita corregir errores y actualizar los datos del modelo.

## 2.2 Aprendizaje Máquina

El aprendizaje máquina (*machine Learning*) es un subcampo de la inteligencia artificial que permite a las computadoras aprender a realizar tareas sin ser programadas explícitamente para cada una de ellas. En lugar de seguir instrucciones paso a paso, los sistemas de aprendizaje máquina utilizan algoritmos

para identificar patrones en grandes conjuntos de datos y hacer predicciones o tomar decisiones basadas en esa información [13].

Existen diferentes tipos de aprendizaje máquina [14]:

- a. Aprendizaje supervisado: la computadora recibe datos etiquetados, es decir, se le indica cuál es la respuesta correcta para cada ejemplo. Por ejemplo, determinar si un correo es correo basura (*spam*) o no.
- b. Aprendizaje no supervisado: la computadora analiza datos sin etiquetas y busca patrones ocultos por sí misma. Por ejemplo, agrupar a los clientes de una tienda en diferentes segmentos basados en sus hábitos de compra.
- c. Aprendizaje por refuerzo: la computadora aprende a tomar decisiones en un entorno interactivo, a partir de recompensas o penalizaciones según los resultados de sus acciones. Por ejemplo, un programa de ajedrez podría aprender a jugar mejor a través de millones de partidas contra sí mismo.

Esta investigación se desarrolló a partir del aprendizaje supervisado. Este tipo de aprendizaje puede ser de dos clases [15]:

- a. Por clasificación: se utiliza cuando la variable objetivo que queremos predecir es categórica, es decir, pertenece a un conjunto finito de clases o etiquetas. Los algoritmos comunes son: árboles de decisión, bosques aleatorios, Bayes ingenuo, redes neuronales, máquina de soporte vectorial, vecino más próximo entre otros. Esta clase es la que empleamos en esta investigación.
- b. Por regresión: se usa cuando la variable objetivo que queremos predecir es numérica. Los algoritmos o métodos comunes son: regresión lineal, logística y de árboles de decisión.

El aprendizaje máquina tiene un amplio rango de aplicaciones en diversas áreas: [11]: reconocimiento de imágenes, procesamiento del lenguaje natural, recomendaciones, detección de fraudes, vehículos autónomos y educación. Esta investigación se enmarca en la última área, pues busca predecir el nivel de distracción de los estudiantes por el uso de redes sociales.

### 2.3 Métricas de clasificación

Las métricas típicas de recuperación de información son precisión y exhaustividad [16]. La precisión (Ecuación 1) es el número de documentos correctamente clasificados sobre el total de documentos clasificados, para una categoría dada. La exhaustividad (Ecuación 2) es el número de documentos clasificados correctamente sobre el total de documentos de esa categoría.

$$prec = \frac{|\{obj_{rel}\} \cap \{obj_{rec}\}|}{obj_{rec}} \quad (1)$$

$$exh = \frac{|\{obj_{rel}\} \cap \{obj_{rec}\}|}{obj_{rec}} \quad (2)$$

Donde,  $obj_{rel}$  corresponde a los objetos relevantes,  $obj_{rec}$  corresponde a los objetos recuperados.

$$F_{\beta} = (1 + \beta^2) * \frac{(prec * exh)}{obj_{rel}} \quad (3)$$

La medida F1 (Ecuación 3) es una media armónica ponderada de exhaustividad ( $exh$ ) y precisión ( $prec$ ); donde,  $\beta$  es una variable para dar preferencia a la exhaustividad o a la precisión. Cuando  $\beta > 1$  entonces la preferencia es para la precisión, cuando  $\beta < 1$  entonces la preferencia es para la exhaustividad.

## 2.4 Herramientas y librerías

La investigación empleó diversas herramientas para el tratamiento de la información y los algoritmos predictivos:

- a. *Google forms*: la herramienta de formularios en línea de Google [17] fue empleada para obtener la información de forma virtual; se elaboró una encuesta y fue distribuida a través de las redes sociales para ser contestada.
- b. SPSS: la información recopilada fue tratada con el software estadístico SPSS [18]; se llevaron a cabo las fases del proceso de minería: selección, limpieza transformación.
- c. SciKitLearn: biblioteca de código abierto y gratuita para el lenguaje de programación Python, especializada en aprendizaje automático [19]; los algoritmos empleados en esta investigación fueron tomados de esta biblioteca.

## Metodología

### 3.1 Integración y recopilación

Para la integración y recopilación de los datos se diseñó un instrumento tipo encuesta [20]. Este instrumento contiene preguntas relacionadas con el uso de las redes sociales y se aplicó a estudiantes universitarios.

La población estuvo constituida por alumnos de Universidades de la zona norte del estado de Sinaloa, principalmente estudiantes de la Universidad Autónoma de Sinaloa (UAS) en la Facultad de Ingeniería Mochis y de la Universidad Autónoma Indígena de México (UAIM) cuya participación estuvo a cargo de la Unidad Virtual. También participaron estudiantes de otras universidades, siendo estas una minoría.

En la UAS se aplicaron las encuestas en la Facultad de Ingeniería Mochis en cuatro carreras: Ingeniería Civil, Ingeniería Geodésica, Ingeniería de Software e Ingeniería en Procesos Industriales.

La Unidad Virtual consideró las siguientes carreras: Licenciatura en Derecho, Licenciatura en Psicología, Licenciatura en Educación Preescolar, Ingeniería en Biotecnología.

### 3.2 Instrumento aplicado

La encuesta con preguntas cerradas y algunas abiertas fue el instrumento para recolectar información de los estudiantes. Las secciones con sus preguntas y respuestas son las siguientes:

1. Datos generales:

- a. Género (hombre, mujer, prefiero no decirlo)
- b. Edad (entero numérico)
- c. Nombre de la Universidad donde estudias (Cadena)
- d. Nombre de la carrera que estudias (Cadena)
- e. Grado actual (año escolar) (numérico entero)
- f. Promedio aproximado a la fecha (real entre 0 y 10).

2. Aspectos generales de uso de las redes sociales

- a. Tiempo de uso de las redes sociales en horas diarias (entero de 0 a 24)
- b. ¿Con cuál tipo de dispositivos te conectas principalmente a las redes sociales? (Celular, laptop, PC, Tablet, otro)
- c. ¿Cuáles redes sociales usas? (YouTube, X, Instagram, WhatsApp, Facebook, TikTok, Telegram, otros)
- d. ¿Cuál red social es tu favorita? (YouTube, X, Instagram, WhatsApp, Facebook, TikTok, Telegram, otros)
- e. ¿Dónde usas principalmente las redes sociales? (Escuela, hogar, sitio público, transporte público, *cyber*, trabajo, otro)
- f. ¿Para qué usas principalmente las redes sociales? (ver videos, publicar videos, publicar y ver fotos, publicar texto, comentar otras publicaciones, chatear, compra/venta, solo ver publicaciones, otro)

### 3. Las redes sociales en la educación

- a. En qué grado usas las redes sociales para el ambiente educativo (nada, poco, regular, mucho, demasiado)
- b. Considerando el aspecto educativo, ¿cuál red social usas principalmente para el intercambio de documentos? (YouTube, X, Instagram, WhatsApp, Facebook, TikTok, Telegram, otros)
- c. Considerando el aspecto educativo, ¿cuál red social usas principalmente para comunicación (textual)? (YouTube, X, Instagram, WhatsApp, Facebook, TikTok, Telegram, otros)
- d. Considerando el aspecto educativo, ¿cuál red social usas principalmente para organización del trabajo en equipo? (YouTube, X, Instagram, WhatsApp, Facebook, TikTok, Telegram, otros)
- e. Considerando el aspecto educativo, ¿cuál red social usas principalmente para publicar videos educativos? (YouTube, X, Instagram, WhatsApp, Facebook, TikTok, Telegram, otros)
- f. ¿Has tenido algún problema en el ámbito educativo por el uso de redes sociales? (sí, no)
- g. ¿Qué tan frecuentemente te distraes de tus labores académicas por el uso de las redes sociales? (Nunca, casi nunca, regularmente, frecuentemente, muy frecuentemente)
- h. ¿Consideras que las redes sociales son útiles para fines educativos? (sí, no, tal vez).

Las variables empleadas para las predicciones comprenden datos generales de los estudiantes relacionados con tiempo de uso de redes sociales y las plataformas utilizadas todo ello en el contexto educativo. Este conjunto de variables permite identificar malos hábitos de los estudiantes y su impacto en el rendimiento: se espera que el uso prolongado de plataformas sociales incida negativamente en las calificaciones de los estudiantes y, a su vez, que permita determinar qué redes sociales se asocian con los promedios académicos más bajos.

### 3.3 Selección, limpieza y transformación

#### 1. Selección

El instrumento recopiló 1208 encuestas contestadas; sin embargo, se detectaron respuestas incoherentes en algunas encuestas. Tras su depuración, quedaron 1185 encuestas útiles.

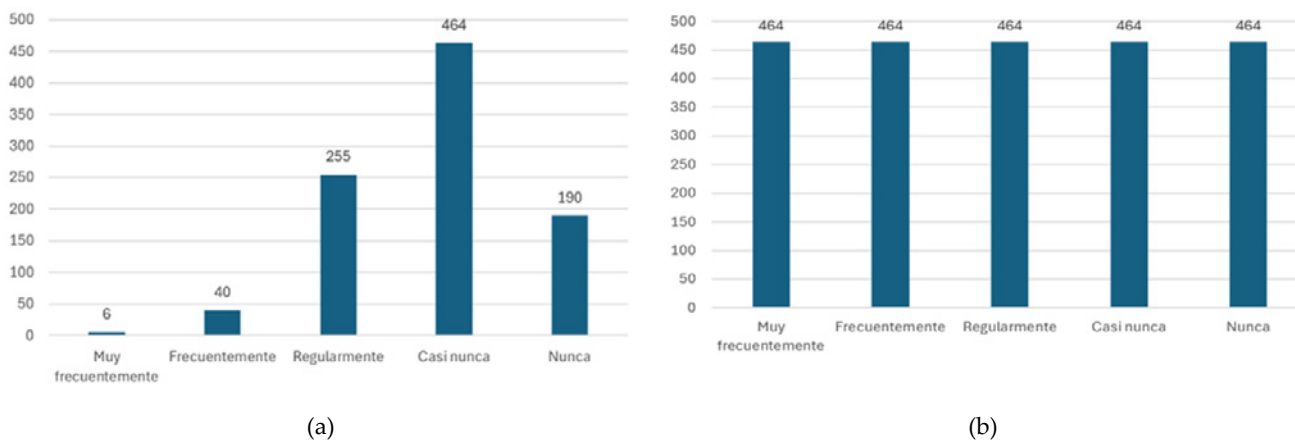
## 2. Limpieza

Desde la recopilación, se encontraron datos anómalos en distintas variables tipo escala, como el tiempo de uso de redes sociales, la edad o el promedio, por lo que se decidió eliminar esas instancias.

## 3. Transformación

Para evitar que variables sin relacionar afectaran los resultados de los algoritmos, se decidió elegir aquellas cuya distribución de frecuencias fuera variada, es decir, que las respuestas a las preguntas no estuviesen en su mayoría recargadas en una sola opción. Por lo tanto, se seleccionaron las siguientes variables: género, edad, promedio aproximado a la fecha, tiempo de uso de las redes sociales en horas diarias, ¿para qué usas principalmente las redes sociales?, ¿en qué grado usas las redes sociales para el ambiente educativo?, considerando el aspecto educativo, ¿cuál red social usas principalmente para publicar videos educativos?, ¿qué tan frecuente te distraes de tus labores académicas por el uso de las redes sociales? (Clase)

Además, fue necesario hacer un balance de clases desproporcionadas. Para ello se creó un nuevo atributo y se eliminó el previo (ver Figura 1). Se eligió igualar las clases a la de mayor frecuencia, en este caso 464. Se empleó el método SMOTE (Synthetic Minority Over-sampling Technique) para el balanceo de clases. Este método no duplica instancias existentes, sino que crea nuevas instancias sintéticas de la clase minoritaria que están cerca de las instancias originales en el espacio de características. Esto añade diversidad sin duplicación directa [21][21].

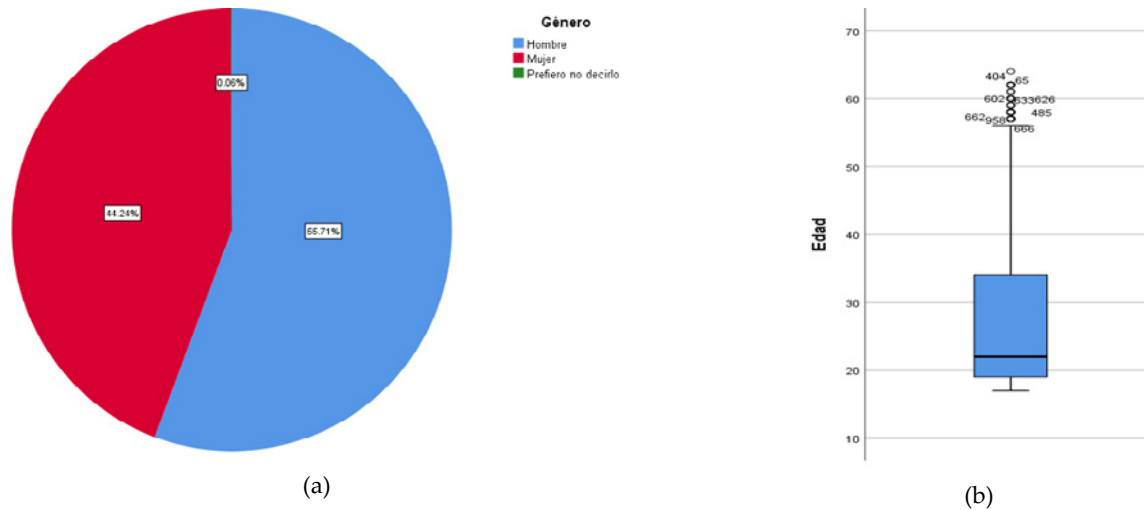


**Figura 1.** Atributo de clase distracción por redes sociales antes (a) y después (b) del balanceo.

**Fuente:** elaboración propia.

### 3.4 Análisis descriptivo

A continuación, se muestra un análisis descriptivo de los atributos seleccionados para la predicción. De acuerdo con el atributo *género*, se muestra un mayor porcentaje de hombres que de mujeres, y un bajo porcentaje de personas que prefieren no decir su género (ver Figura 2 a).



**Figura 2.** (a) Porcentaje del atributo género. (b) Distribución de la variable edad.

**Fuente:** elaboración propia.

La media de edad obtenida es de 27.08 con una desviación estándar de 10.33. La Figura 2 b muestra la distribución de los datos con bastantes personas entre 17 y 23 años, además se muestran varias personas mayores a 55 años todavía estudiando.

Los estudiantes en general muestran un promedio de 8.81 en sus calificaciones con una desviación estándar de 0.96. En general se muestran buenas calificaciones, aunque hay algunas muy bajas (ver Figura 3 a).

El tiempo de uso de las redes sociales es de 5.05 horas en promedio, por lo que se aprecia una distribución sesgada a la izquierda en la Figura 3 b. Hay varios estudiantes que exageran en sus estimaciones al poner cifras cercanas a las 24 horas.

La mayor parte de los estudiantes usan las redes sociales para ver videos, chatear o solo ver publicaciones. La Figura 4 a muestra el resto de los porcentajes de otras actividades.

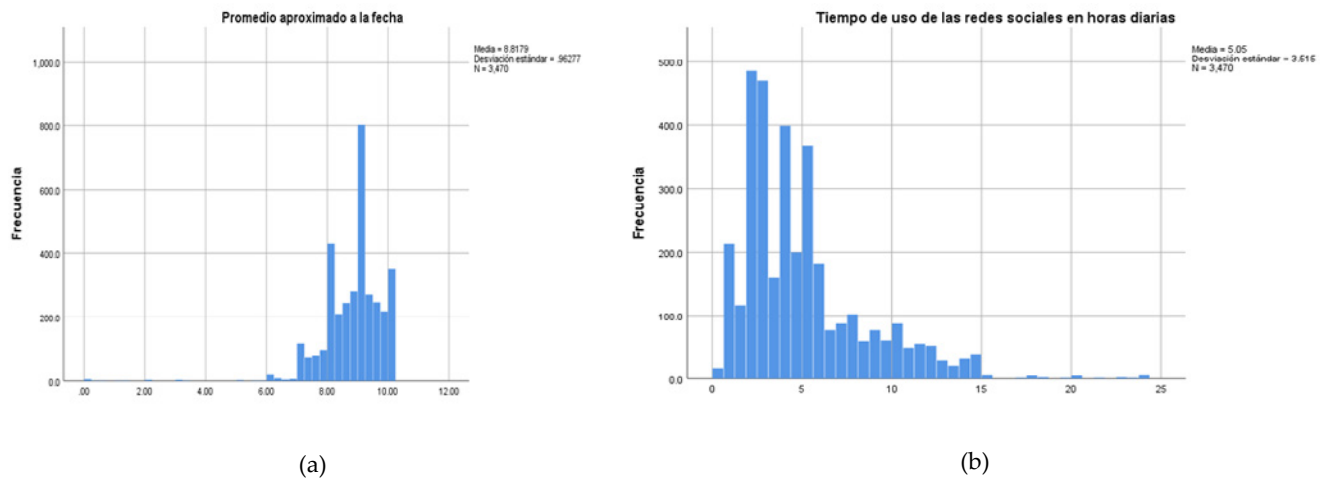


Figura 3. Histograma del promedio (a) y del tiempo de uso de las redes sociales en horas diarias (b).

Fuente: elaboración propia.

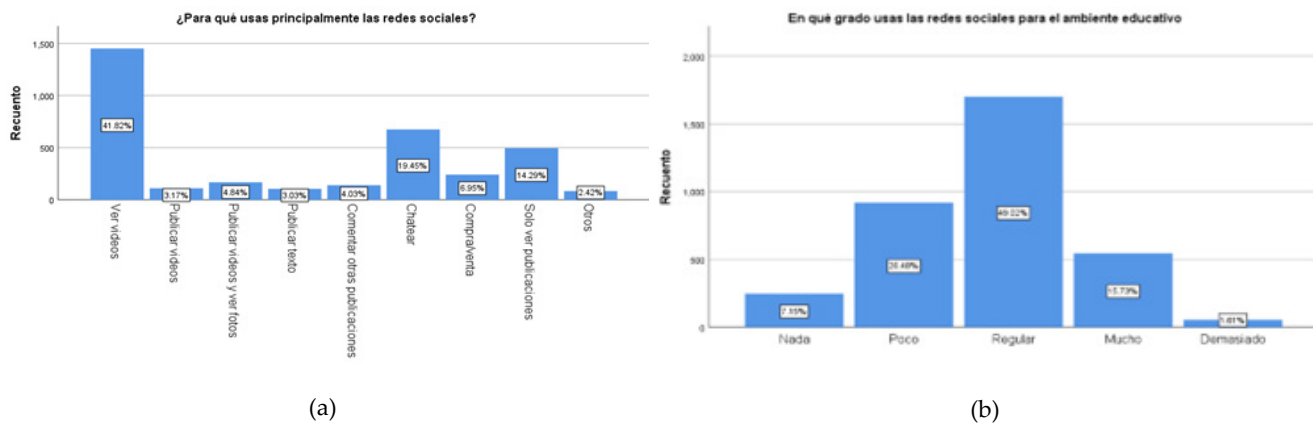
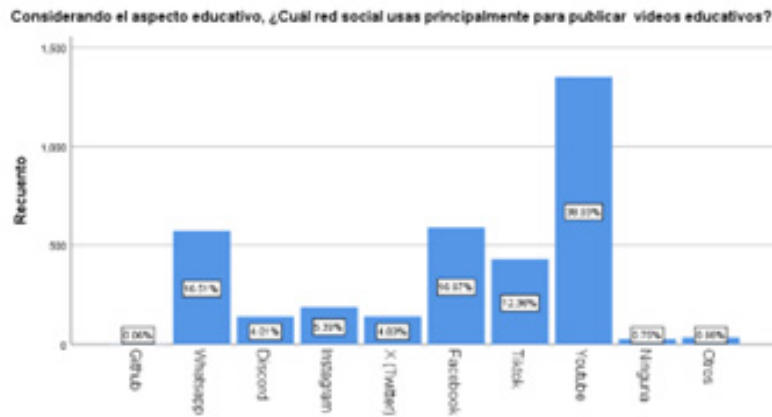


Figura 4. (a) ¿Para qué usas principalmente las redes sociales?  
(b) ¿Para qué usas las redes sociales para el ambiente educativo?

Fuente: elaboración propia.

Por otra parte, los usuarios usan de forma regular las redes sociales para el ambiente educativo, siendo la distribución un poco simétrica hacia ambos lados, aunque ligeramente cargada hacia poco o nada de uso (ver Figura 4 b).

Finalmente, la Figura 5 muestra cual red usan los estudiantes principalmente para publicar videos educativos, destacando YouTube, Facebook y WhatsApp en ese orden.



**Figura 5.** ¿Cuál red social usas principalmente para publicar videos educativos?

**Fuente:** elaboración propia.

La distribución del atributo de clase ¿Qué tan frecuente te distraes de tus labores académicas por el uso de las redes sociales? Se muestra en la Figura 1, dado el balanceo de cada categoría de la clase todas quedaron con la misma frecuencia (464).

### 3.5 Minería de datos

Existen diferentes tareas de la minería de datos tales como clasificación, regresión, agrupamiento, correlación y reglas de asociación [11]. Esta investigación se enfoca en la clasificación, es decir, predecir el nivel de distracción (muy frecuente, frecuente, regularmente, casi nunca, nunca) que puede tener un estudiante en el uso de las redes sociales.

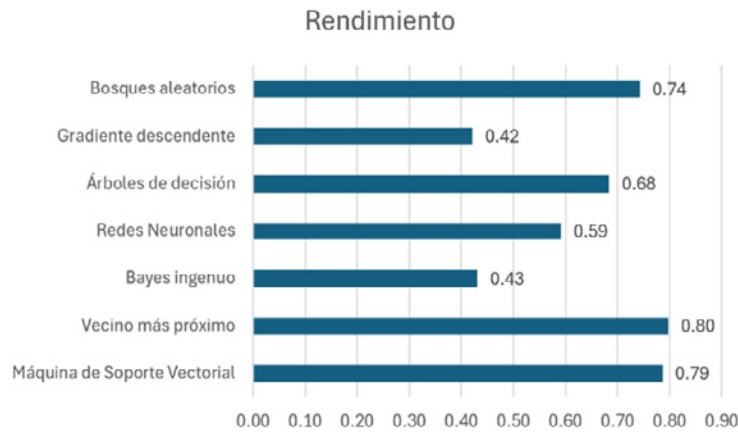
Para elegir la técnica de minería apropiada para los datos, se hace un comparativo con 7 algoritmos de clasificación: máquina de soporte vectorial, vecino más próximo, Bayes ingenuo, redes neuronales, árboles de decisión, gradiente descendente y bosques aleatorios.

### 3.6 Evaluación e interpretación

El proceso de evaluación incluye las siguientes etapas: 1) importar los datos, 2) seleccionar los atributos importantes, 3) aplicar un balanceo de clases, 4) seleccionar los clasificadores, 5) dividir el conjunto de datos en 80 % entrenamiento y 20 % de prueba, 6) definir los parámetros de configuración del clasificador, 7) entrenar el algoritmo con una validación cruzada de 5 pliegues, 8) mostrar la mejor configuración y el mejor resultado.

El último punto necesitó de varias iteraciones de prueba seleccionando los mejores clasificadores ya que tenían un rendimiento muy parecido.

Los resultados de los datos de entrenamiento son los que se muestran en la Figura 6. Tal como se aprecia, el vecino más próximo (0.80), la máquina de soporte vectorial (0.79) y bosques aleatorios (0.74) son los que muestran mejor rendimiento con los datos de prueba.



**Figura 6.** Rendimiento de los algoritmos de clasificación.

**Fuente:** elaboración propia.

Con los tres mejores algoritmos se midió de nuevo el rendimiento empleando también 5 pliegues, pero solo utilizando la mejor configuración obtenida con GridSearchCV. Los resultados se aprecian en la Tabla 1 en la cual se destaca el rendimiento similar de la máquina de soporte vectorial y el vecino más próximo.

**Tabla 1.** Resultados de los tres mejores algoritmos

Algoritmo	Iteraciones					Prom.
	1	2	3	4	5	
Máquina de Soporte Vectorial	0.7435	0.8026	0.8401	0.8487	0.8141	0.8098
Vecino más próximo	0.7550	0.7954	0.8386	0.8271	0.8314	0.8095
Bosques aleatorios	0.6138	0.6888	0.7695	0.7666	0.7622	0.7202

**Fuente:** elaboración propia.

Una tercera prueba se consideró solo para los algoritmos de la máquina de soporte vectorial y el vecino más próximo; por lo cual se realizó una validación simple con 80 % de entrenamiento y 20 % de prueba y los mejores parámetros obtenidos con GridSearchCV. Se obtuvieron la matriz de confusión normalizada y sin normalizar, además de las métricas de precisión, exhaustividad y la medida f1.

Las Tablas 2 y 4 muestran los resultados de la matriz de confusión, en las que columnas y renglones representan las clases, su intersección, su frecuencia de acierto y el porcentaje asociado al total de elementos (139). Además, se representa una escala de colores donde el color rojo representa valores cercanos al cero y el verde valores próximos a uno.

**Tabla 2.** Matriz de confusión para el algoritmo del vecino más próximo

	Muy frec.	Frec.	Reg.	Casi nunca	Nunca
Muy Frec.	138 (0.99)	0 (0.00)	0 (0.00)	1 (0.01)	0 (0.00)
Frec.	1 (0.01)	133 (0.96)	1 (0.01)	3 (0.02)	1 (0.01)
Regularmente	3 (0.02)	4 (0.03)	106 (0.76)	18 (0.13)	8 (0.06)
Casi nunca	2 (0.01)	17 (0.12)	27(0.19)	69(0.5)	24(0.17)
Nunca	0 (0.00)	2(0.01)	7(0.05)	16(0.12)	113(0.82)

Fuente: Autores.

Las Tablas 3 y 5 representan los resultados de las métricas: precisión, exhaustividad y la medida f1 considerando las clases en las columnas. La escala de colores se representa de la misma forma que las tablas 1 y 2.

**Tabla 3.** Métricas de rendimiento para el algoritmo del vecino más próximo

	Muy frec.	Frec.	Reg.	Casi nunca	Nunca
Precisión	0.96	0.85	0.75	0.64	0.77
Exhaustividad	0.99	0.96	0.76	0.5	0.82
Medida f1	0.98	0.9	0.76	0.56	0.8
Soporte	139	139	139	139	138

Fuente: Autores.

El promedio de los rendimientos para el algoritmo del vecino más próximo es: precisión es igual a 0.79, exhaustividad es igual a 0.81 y la medida f1 es igual 0.80. Mientras que la máquina de soporte vectorial tuvo una precisión igual 0.79, una exhaustividad de 0.81 y una medida f1 de 0.80.

**Tabla 4.** Matriz de confusión para el algoritmo máquina de soporte vectorial

	Muy frec.	Frec.	Reg.	Casi nunca	Nunca
Muy Frec.	139 (1.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)
Frec.	0 (0.00)	134 (0.96)	2 (0.01)	2 (0.01)	1 (0.01)
Regularmente	1 (0.01)	10 (0.07)	99 (0.71)	19 (0.14)	10 (0.07)
Casi nunca	1 (0.01)	11 (0.08)	27 (0.19)	83 (0.60)	17 (0.12)
Nunca	0 (0.00)	6 (0.04)	9 (0.07)	15 (0.11)	108 (0.78)

Fuente: Autores.

**Tabla 5.** Métricas de rendimiento para el algoritmo máquina de soporte vectorial

	Muy frec.	Frec.	Reg.	Casi nunca	Nunca
Precisión	0.95	0.85	0.75	0.64	0.77
Exhaustividad	0.99	0.96	0.76	0.5	0.82
Medida f1	0.98	0.9	0.76	0.56	0.8

Fuente: Autores.

### 3.7 Difusión, uso y monitorización

Con el modelo ya entrenado, se podrían realizar recomendaciones a algunos estudiantes respecto al uso inadecuado de plataformas sociales. Dichas recomendaciones pueden servir como insumos para que el estudiante cree conciencia sobre el mal uso de las redes sociales. Cada alumno al final tomará la decisión sobre sus acciones; sin embargo, algunos tomarán las decisiones correctas, lo cual busca esta investigación.

El modelo de minería debe monitorearse continuamente, puesto que los datos cambian constantemente y los patrones que eran válidos en el pasado pueden dejar de serlo. El monitoreo permite identificar estos cambios y ajustar el modelo oportunamente. Además, posibilita medir la precisión, la sensibilidad y otras métricas clave para verificar que el modelo sigue cumpliendo los objetivos establecidos. De esta forma, el monitoreo puede detectar valores atípicos o comportamientos inusuales en los datos que podrían indicar problemas en el modelo o en los datos fuente.

## Discusión

La distracción de los estudiantes a causa del uso de las redes sociales es un tema de creciente preocupación en el ámbito educativo. Las investigaciones indican que el uso excesivo de aplicaciones sociales afecta negativamente el rendimiento académico de los estudiantes al ser sometidos a distracciones.

El uso excesivo de comunidades virtuales, como Facebook, puede causar distracción cognitiva, lo que a su vez afecta el rendimiento académico de los estudiantes [22]. Los estudiantes que pasan más tiempo en redes sociales tienden a tener un menor rendimiento académico debido a la distracción [23], [24]. Además, las características tecnológicas como la socialización, las comparaciones sociales, el disfrute y la búsqueda de información son determinantes del uso excesivo de plataformas sociales, lo que lleva a distracciones [22]. Además, el miedo a perderse algo también impulsa el uso de sitios de interacción social durante las clases [25].

Por otra parte, el uso de redes sociales no solo afecta el rendimiento académico, sino que también puede llevar a una pérdida de interés en los estudios, interrupciones del comportamiento y barreras

para el auto-mejoramiento [26]. Aunado a esto, puede contribuir a problemas de salud mental como ansiedad y depresión [24].

Se ha identificado que el control Cognitivo-Conductual (CC) puede mitigar los efectos negativos del uso excesivo de redes sociales en la distracción cognitiva [22]. Es decir, el control CC se centra en identificar y cambiar patrones de pensamiento negativos y comportamientos poco saludables. Al comprender cómo nuestros pensamientos afectan nuestras emociones y acciones, podemos desarrollar estrategias para modificarlos y mejorar nuestra calidad de vida.

Dada la influencia perjudicial de las redes sociales en el ámbito educativo, se hace imperativo que las instituciones diseñen políticas que ayuden a los estudiantes a gestionar su uso de forma más responsable y productiva [24]. La predicción proactiva de las tendencias de mal uso se revela, por tanto, como una herramienta invaluable para el diseño de dichas estrategias.

## Conclusiones

Esta investigación, en primera instancia, construyó un conjunto de datos a través de un instrumento para recopilar información sobre el uso de redes sociales en el ambiente educativo. Posteriormente, aplicó un proceso de minería de datos para encontrar patrones predictivos con el fin de determinar el nivel de distracción que generan las plataformas sociales en el aprendizaje.

Los resultados indican que podemos predecir el nivel de distracción de los estudiantes considerando la escala: muy frecuente, frecuente, regular, casi nunca y nunca. Esta predicción se da en 4 de cada 5 casos, es decir con un 80 % de efectividad. Los algoritmos que mejor rendimiento tuvieron fueron la máquina de soporte vectorial y el algoritmo del vecino más próximo.

Las redes sociales, aunque útiles para la comunicación y el aprendizaje colaborativo, pueden ser una fuente significativa de distracción para los estudiantes, lo cual afecta su rendimiento académico y bienestar mental. Es esencial que tanto los estudiantes como los educadores sean conscientes de estos efectos y trabajen juntos para desarrollar estrategias que minimicen las distracciones y promuevan un uso más equilibrado de las comunidades sociales.

El modelo predictivo mostrado en esta investigación puede ayudar a identificar a los estudiantes en riesgo de distracción alta al inicio de las asignaturas materia, al caracterizar su uso de las redes sociales. Esto permitiría una intervención del profesor antes de que el problema afecte seriamente su aprendizaje.

Como trabajo futuro, se podría reestructurar el conjunto de datos agregando nuevas preguntas y quitando otras. Esto puede subsanar algunas limitaciones que se presentaron en esta investigación: por ejemplo, hubo algunas preguntas que no tuvieron el impacto deseado y tuvieron que ser eliminadas en

las instancias de predicción dado su escaso aporte al proceso de aprendizaje máquina. De esta forma, se podría incrementar el rendimiento de la predicción y, de esa manera, lograr un mayor impacto en las recomendaciones. Además, aún queda pendiente el desarrollo de una herramienta de software que cuente con una interfaz y pueda brindar apoyo a departamentos de las universidades con la detección de mal uso de las redes sociales.

## Referencias

- [1] Y. Venugeetha, R. Ranjeet, y K. Rishabh, "Social Networking Sites in Daily Life: Benefits and Threats," en *Artificial Intelligence and Communication Technologies*. India: SCRS, 2022, pp. 51–64. doi: <https://doi.org/10.52458/978-81-955020-5-9-5>
- [2] T. Sanwal, S. Yadav, S. Avasthi, A. Prakash, y M. Tyagi, "Social Media and Networking Applications in the Education Sector," en *2023 2nd Edition of IEEE Delhi Section Flagship Conference (DELCON)*. IEEE, feb. 2023, pp. 1–6. doi: <https://doi.org/10.1109/DELCON57910.2023.10127547>
- [3] I. Milošević, D. Živković, S. Arsić, y D. Manasijević, "Facebook as virtual classroom - Social networking in learning and teaching among Serbian students," *Telematics and Informatics*, vol. 32, núm. 4, pp. 576–585, 2015. doi: <https://doi.org/10.1016/j.tele.2015.02.003>
- [4] M. R. Buxarrais, "Redes sociales y educación," *Education in the Knowledge Society (EKS)*, vol. 17, núm. 2, pp. 15–20, jul. 2016. doi: <https://doi.org/10.14201/eks20161721520>
- [5] G. A. Flores Lagla, J. C. Chancusig Chisag, J. A. Cadena Moreano, O. A. Guaypatín Pico, y R. H. Montaluisa Pulloquina, "La influencia de las redes sociales en los estudiantes universitarios," *Boletín Redipe*, vol. 6, núm. 4, pp. 56–64, oct. 2017.
- [6] J. Chen, "Social Media Addiction and Consequences in Adolescents," *Lecture Notes in Education Psychology and Public Media*, vol. 7, núm. 1, pp. 291–296, may. 2023. doi: <https://doi.org/10.54254/2753-7048/7/20220823>
- [7] Q. Du, "The Negative Impact of Social Media on Adolescents," *Lecture Notes in Education Psychology and Public Media*, vol. 6, núm. 1, pp. 305–309, may. 2023. doi: <https://doi.org/10.54254/2753-7048/6/20220346>
- [8] H. Pang, "The Negative Impact of Social Media on People's Lives," en *Proceedings of the 2021 International Conference on Social Development and Media Communication (SDMC 2021)*. Atlantis Press, 2022, pp. 554–557. doi: <https://doi.org/10.2991/assehr.k.220105.102>
- [9] W. Frawley, G. Piatetsky-Shapiro, y C. Matheus, "Knowledge Discovery in Databases: An overview," *AI Magazine*, vol. 13, núm. 1, pp. 57–70, 1992. <https://ojs.aaai.org/aimagazine/index.php/aimagazine/article/download/1011/929>
- [10] U. Fayyad, G. Piatetsky-Shapiro, y P. Smyth, "From data mining to knowledge discovery in databases," *AI Magazine*, vol. 17, núm. 3, pp. 37–53, 1996.
- [11] J. Hernández Orallo, Ma. J. Ramírez Quintana, y C. Ferri Ramírez, *Introducción a la minería de datos*. Madrid: Pearson Educación, 2004.
- [12] S. R. Timarán-Pereira, I. Hernández-Arteaga, S. J. Caicedo-Zambrano, y A. Hidalgo-Troya, "El proceso de descubrimiento de conocimiento en bases de datos," en *Descubrimiento de patrones de desempeño académico con*

*árboles de decisión en las competencias genéricas de la formación profesional*. Bogotá, Colombia: Universidad Cooperativa de Colombia, 2016, pp. 63–86. doi: <https://doi.org/10.16925/9789587600490>

- [13] R. Thomas y R. Gupta, "A Survey on Machine Learning Approaches and Its Techniques," en *2020 IEEE International Students' Conference on Electrical, Electronics and Computer Science (SCEECS)*. IEEE, 2020, pp. 1–6. doi: <https://doi.org/10.1109/SCEECS48394.2020.190>
- [14] A. Sharma, A. Kaur, y A. Semwal, "Supervised and Unsupervised Prediction Application of Machine Learning," en *2022 International Conference on Cyber Resilience (ICCR)*. IEEE, 2022, pp. 1–5. doi: <https://doi.org/10.1109/ICCR56254.2022.9996063>
- [15] V. Gupta, V. K. Mishra, P. Singhal, y A. Kumar, "An Overview of Supervised Machine Learning Algorithm," en *2022 11th International Conference on System Modeling & Advancement in Research Trends (SMART)*. IEEE, 2022, pp. 87–92. doi: <https://doi.org/10.1109/SMART55829.2022.10047618>
- [16] A. Nenkova y K. McKeown, "Automatic Summarization," *Foundations and Trends in Information Retrieval*, vol. 5, núm. 2–3, pp. 103–233, 2011. doi: <https://doi.org/10.1561/1500000015>
- [17] Google, "Google Forms," Google LLC. Consultado: el 8 de mar. de 2024. [En línea]. Disponible en: [https://www.google.com/intl/es\\_mx/forms/about/](https://www.google.com/intl/es_mx/forms/about/)
- [18] IBM Corp., *IBM SPSS Statistics for Windows*, versión 28.0. Armonk, NY: IBM Corp., 2021.
- [19] F. Pedregosa *et al.*, "Scikit-learn: Machine Learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.
- [20] E. Cowles y E. Nelson, *An Introduction to Survey Research*. New York: Business Expert Press, 2015.
- [21] O. Loyola-González, J. F. Martínez-Trinidad, J. A. Carrasco-Ochoa, y M. García-Borroto, "Study of the impact of resampling methods for contrast pattern based classifiers in imbalanced databases," *Neurocomputing*, vol. 175, pp. 935–947, 2016. doi: <https://doi.org/10.1016/j.neucom.2015.04.120>
- [22] A. Masood, A. Luqman, Y. Feng, y A. Ali, "Adverse consequences of excessive social networking site use on academic performance: Explaining underlying mechanism from stress perspective," *Computers in Human Behavior*, vol. 113, art. 106476, 2020. doi: <https://doi.org/10.1016/j.chb.2020.106476>
- [23] M. Kolhar, R. N. A. Kazi, y A. Alameen, "Effect of social media use on learning, social interactions, and sleep duration among university students," *Saudi Journal of Biological Sciences*, vol. 28, núm. 4, pp. 2216–2222, 2021. doi: <https://doi.org/10.1016/j.sjbs.2021.01.010>
- [24] N. Andleeb, S. Rafique, y Quratulain, "Use of Social Media Tools by Undergraduates: Students and Teachers' Perspective," *Journal of Social Sciences Review*, vol. 3, núm. 1, pp. 758–767, mar. 2023. doi: <https://doi.org/10.54183/jssr.v3i1.218>
- [25] C. Shane-Simpson y T. Bakken, "Students' fear of missing out predicts in-class social media use," *Teaching of Psychology*, vol. 51, núm. 2, pp. 141–150, 2024. doi: <https://doi.org/10.1177/00986283211060752>
- [26] M. O. Sagabala, S. M. A. Regidor, M. C. M. Dela Torre, y A. M. M. Monteza, "A Qualitative Inquiry on the Behavior of English Major Students Towards Social Media Distractions," *European Journal of Education Studies*, vol. 10, núm. 8, 2023. doi: <https://doi.org/10.46827/ejes.v10i8.4907>

