

PREDICCIÓN DE ADICCIÓN EN ESTUDIANTES A LAS REDES SOCIALES

Yuliani Saray Beltrán Nájera & Melani Juliana Tirado Salas

Universidad del Magdalena-Diplomado de Analítica de Datos para la Toma de Decisiones.

Santa Marta, Colombia.

2025

yulianibeltransn@unimagdalena.edu.co; mtirado@unimagdalena.edu.co

1. RESUMEN EJECUTIVO

El objetivo de este análisis es clasificar a los estudiantes como adictos o no adictos a las redes sociales, identificando las variables que influyen en este comportamiento. Para ello, se utilizó un conjunto de datos que contiene 705 registros de estudiantes entre los 16 y 25 años, provenientes de diferentes países. Las variables analizadas incluyen edad, género, nivel académico, país de residencia, promedio de horas diarias en redes sociales, plataforma más utilizada, percepción del impacto en el rendimiento académico, horas de sueño, estado de la salud mental autoevaluada, estado sentimental y conflictos ocasionados por el uso de redes sociales. Esta información fue recopilada a través de una encuesta en línea aplicada en el primer trimestre del año 2025, basada en escalas como la Escala de Adicción a Redes Sociales de Bergen.

Los hallazgos más relevantes del modelo indican que existen combinaciones de factores que se repiten en los perfiles de estudiantes clasificados como adictos. Por ejemplo, aquellos que consideran que las redes sociales afectan negativamente su rendimiento académico, que duermen menos de 7.15 horas por noche y que tienen menos de 21 años, presentan una mayor probabilidad de ser clasificados como adictos. Además, el uso frecuente de plataformas como Instagram o TikTok, una salud mental autoevaluada por debajo de 6.5 puntos y ciertos estados sentimentales como estar soltero o en una situación complicada también se asocian con patrones adictivos. El modelo reveló reglas específicas que, combinando estas variables, permiten predecir la adicción a redes sociales, destacando así la importancia del análisis multivariable para comprender este fenómeno.

En el proceso de modelado se empleó un árbol de decisión construido mediante el algoritmo Random Forest, por su capacidad para manejar relaciones no lineales y detectar patrones complejos en conjuntos de datos heterogéneos. Este modelo fue entrenado para identificar combinaciones de variables que permiten clasificar a un estudiante como adicto o no, obteniendo métricas de desempeño como precisión (Clase 0=1.0; Clase 1= 0.99), recall (Clase 0= 0.98; Clase 1= 1.0) y F1-score (Clase 0=0.99; Clase 1= 0.99), lo que respalda la validez de los resultados obtenidos.

A partir de estos resultados, se concluye que la adicción a redes sociales no puede explicarse por una sola variable, sino por una interacción entre el tiempo de exposición, el contexto emocional, la etapa de desarrollo del estudiante y las plataformas preferidas. Esta comprensión fundamentada permite no solo anticipar comportamientos de riesgo, sino también diseñar estrategias de prevención y acompañamiento más efectivas para los estudiantes que se encuentren en grupos vulnerables. En este sentido, el estudio contribuye a la toma de decisiones informadas en entornos educativos y de salud mental, promoviendo un uso más saludable de las redes sociales entre los jóvenes.

1.1. OBJETIVO DEL PROYECTO

El presente análisis tiene como principal objetivo responder la pregunta: ¿Pueden identificarse patrones de conducta personales que predigan la adicción a las redes sociales en estudiantes? Con base en esta pregunta, se busca determinar qué variables presentan una relación significativa con los niveles de adicción reportados por los encuestados, permitiendo así establecer perfiles de riesgo.

2. CONTEXTO DEL PÚBLICO OBJETIVO

En la era digital actual, las redes sociales se han convertido en un componente central en la vida de los estudiantes, especialmente aquellos que cursan estudios secundarios, de pregrado o posgrado. Plataformas como Instagram, TikTok y Facebook han redefinido la forma en que los jóvenes se comunican, se informan e incluso desarrollan su identidad personal. Sin embargo, el uso excesivo y desregulado de estas herramientas ha dado lugar a preocupaciones en términos de salud mental, calidad del sueño, rendimiento académico y bienestar emocional. La facilidad de acceso y la constante exposición a contenidos sociales pueden generar comportamientos adictivos que, a largo plazo, comprometen el desarrollo académico y personal de los estudiantes. En este contexto, surge la necesidad de comprender, desde una perspectiva analítica, cuáles son los factores que más inciden en la adicción a las redes

sociales, con el fin de anticipar riesgos, diseñar estrategias de prevención y fomentar el uso responsable de las plataformas digitales.

2.1. DEFINICIÓN DEL PROBLEMA

La pregunta “¿Qué características personales y conductuales predicen la adicción a las redes sociales en estudiantes?” busca explorar las relaciones entre variables como el número de horas de uso diario, la calidad del sueño, el impacto académico, la salud mental autoevaluada y las plataformas preferidas, con la intención de identificar patrones consistentes de comportamiento adictivo. Este problema implica no solo una inquietud en el ámbito educativo y de salud mental, sino también un reto analítico: establecer modelos predictivos confiables que permitan clasificar a los estudiantes como adictos o no adictos, a partir de datos cuantitativos. Comprender cómo influyen estos factores en el desarrollo de conductas adictivas permitirá proponer intervenciones más eficaces en instituciones educativas y entornos familiares, contribuyendo al bienestar integral del estudiante.

3. DESCRIPCIÓN DE LOS DATOS

Los datos utilizados en este análisis provienen de la base de datos Students' Social Media Addiction, disponible públicamente en la plataforma Kaggle y recopilada por Adil Shamim en el año 2025. Este conjunto de datos contiene un total de 705 registros, cada uno correspondiente a la respuesta individual de un estudiante sobre su comportamiento en redes sociales, estado emocional, rendimiento académico y otros factores personales. La información fue recolectada a través de una encuesta en línea aplicada a estudiantes de entre 16 y 25 años, provenientes de distintos niveles académicos (High School, Undergraduate y Graduate) y de países como India, Estados Unidos, Reino Unido, Canadá, Alemania, Japón, entre otros.

Entre las variables relevantes del dataset se encuentran: Student_ID, Age, Gender, Academic_Level, Country, Avg_Daily_Usage_Hours, Most_Used_Platform, Affects_Academic_Performance, Sleep_Hours_Per_Night, Mental_Health_Score, Relationship_Status, Conflicts_Over_Social_Media y Addicted_Score. Estas variables, tanto categóricas como numéricas, permiten realizar un análisis exhaustivo de los patrones que podrían indicar una conducta adictiva frente al uso de redes sociales en el entorno estudiantil.

3.1 ANÁLISIS EXPLORATORIO

En el análisis exploratorio, como primer paso se cargó la base de datos proveniente de Kaggle titulada Students' Social Media Addiction (Adil Shamim, 2025), y se procedió a la verificación de datos faltantes o nulos, sin encontrar ninguno. Esto permitió avanzar directamente con el análisis descriptivo de las variables.

Se elaboró un gráfico boxplot para observar la distribución de las horas de uso promedio diario de redes sociales según si el estudiante es considerado adicto o no. Se evidenció que los estudiantes clasificados como adictos (valor 1 en la variable Addicted) presentan una mediana (5.8 contra 4) de uso diario más altas que aquellos no adictos (valor 0), además se observan valores extremos para la clase adicta de más de 8 horas de uso, lo que refuerza la idea de un patrón intensivo de uso entre quienes presentan conductas adictivas.

También se generó un histograma que muestra la distribución general del uso promedio diario de redes sociales (Avg_Daily_Usage_Hours). La forma del histograma sugiere una distribución ligeramente sesgada hacia la derecha, con la mayoría de los estudiantes concentrados entre 4 y 6 horas de uso diario, aunque se observan casos que superan las 8 horas por día.

Posteriormente, se construyó una matriz de correlación para evaluar la relación entre la variable dependiente y las independientes. Dentro de los hallazgos más destacados, se observó una alta correlación positiva entre la variable Avg_Daily_Usage_Hours de 0.70 y la variable Addicted, lo cual es consistente con la idea de que un mayor tiempo diario de uso se asocia con niveles más altos de adicción. Asimismo, se identificó una fuerte correlación negativa entre Mental_Health_Score y Addicted de -0.84, sugiriendo que los estudiantes con puntuaciones más altas en adicción tienden a reportar un estado de salud mental más bajo. También se detectaron relaciones negativas importantes entre Sleep_Hours_Per_Night y las variables Addicted y Avg_Daily_Usage_Hours de -0.69 y -0.79, indicando que quienes duermen menos tienden a usar más redes sociales y presentar niveles más altos de adicción.

Por último, el análisis exploratorio permitió detectar patrones claros que relacionan el uso intensivo de redes sociales con mayores niveles de adicción, menores horas de sueño y peores puntuaciones de salud mental. Estos hallazgos brindan una base para continuar con análisis más profundos y modelado posterior.

4. PREPARACIÓN Y LIMPIEZA DE DATOS

En el proceso de preparación y limpieza de los datos se creó una nueva variable binaria llamada Addicted, a partir de la variable Addicted_Score. Para ello, se estableció un punto de corte de 7, clasificando como adictos (valor 1) a aquellos estudiantes con una puntuación mayor o igual a ese umbral, y como no adictos (valor 0) a los demás.

Posteriormente, se eliminaron las columnas Student_ID y Addicted_Score debido a que no aportaban valor predictivo directo al modelo. Student_ID es una variable identificadora y Addicted_Score fue reemplazada por la nueva variable binaria.

Las variables categóricas presentes en la base de datos fueron transformadas a variables numéricas utilizando codificación por etiquetas (Label Encoding), lo cual fue necesario para que los modelos de machine learning pudieran procesarlas correctamente. Este procedimiento se aplicó a todas las columnas de tipo objeto. Finalmente, se separó la base de datos en dos conjuntos: uno de características predictoras (X) y otro con la variable objetivo (y), correspondiente a la variable Addicted.

Durante este proceso, se verificó la calidad de la información y no se encontraron valores faltantes, duplicados ni inconsistencias en los datos. La estructura del conjunto de datos es coherente y completa, lo cual permitió avanzar sin necesidad de aplicar técnicas de imputación o corrección adicionales.

5. MODELADO

Para abordar el problema de clasificación sobre la adicción al uso de redes sociales, se optó por utilizar un modelo de clasificación basado en Random Forest, una técnica de ensamblado que combina múltiples árboles de decisión. Esta elección se fundamentó en su capacidad para manejar variables categóricas codificadas, su resistencia al sobreajuste, y su buen desempeño general en problemas con posibles relaciones no lineales entre variables.

Una vez entrenado el modelo, se evaluó su rendimiento utilizando métricas como accuracy, precision, recall y f1-score. A continuación, se resumen los resultados obtenidos en los datos de prueba:

Figura 1. Métricas de rendimiento

Accuracy	0.9929078014184397			
Classification Report:				
	PRECISION	RECALL	F1-SCORE	SUPPORT
0	1.00	0.98	0.99	62
1	0.99	1.00	0.99	79

Fuente: Elaboración propia en Google Colab con datos de Kaggle (Students' Social Media Addiction, Adil Shamim, 2025).

Adicionalmente, se realizó una validación cruzada de 5 folds para verificar la estabilidad del modelo. Los puntajes de accuracy obtenidos en cada fold fueron:

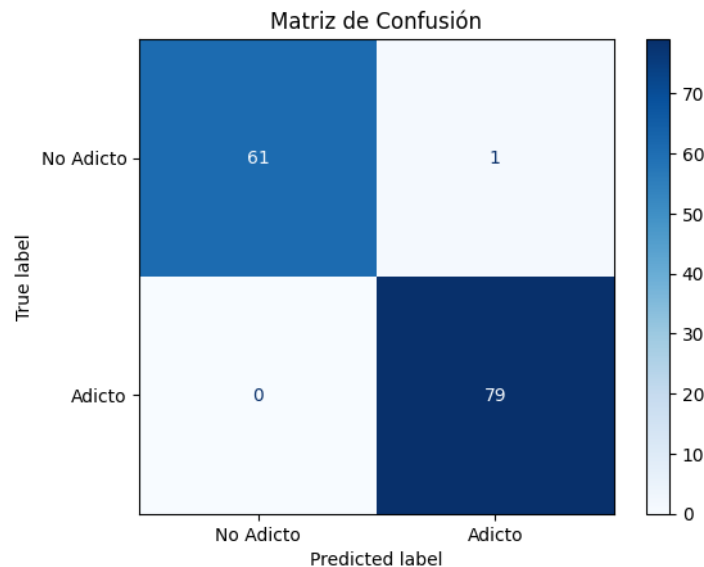
[1.00, 0.95, 0.98, 1.00, 0.99], con un promedio de 0.984 y una desviación estándar de 0.019, lo cual indica un rendimiento sólido y consistente del modelo a través de distintas particiones de los datos.

En términos generales, el modelo cumple satisfactoriamente con el objetivo planteado: identificar de manera precisa a los estudiantes con probabilidad de presentar adicción al uso de redes sociales. Las métricas obtenidas reflejan una alta capacidad predictiva, especialmente destacando el equilibrio entre precision y recall, lo cual es fundamental en contextos donde es importante minimizar tanto los falsos positivos como los falsos negativos.

6. EVALUACIÓN Y RESULTADOS CLAVE

El modelo final alcanzó una precisión de aproximadamente 99.29% y demostró un excelente rendimiento en todas las métricas evaluadas, incluyendo precisión, recall y f1-score para ambas clases (Adicto y No Adicto).

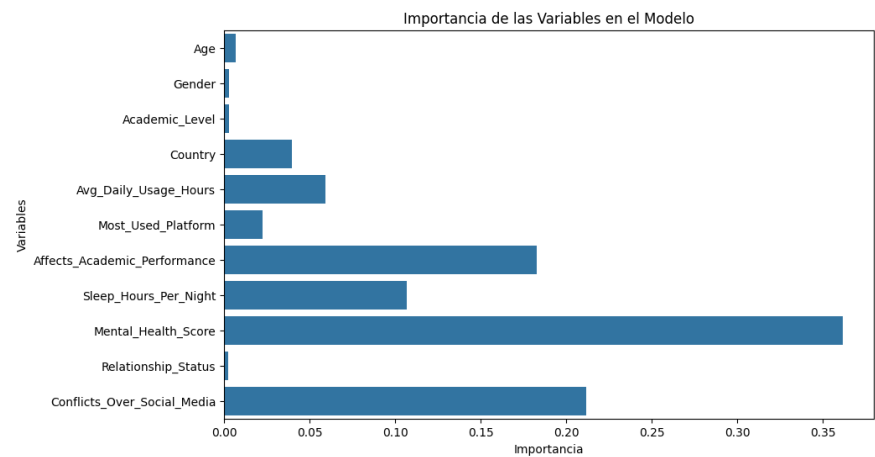
Figura 2. Matriz de Confusión



Fuente: Elaboración propia en Google Colab con datos de Kaggle (Students' Social Media Addiction, Adil Shamim, 2025).

La matriz de confusión muestra un desempeño robusto, con solo una clasificación errónea sobre un total de 141 casos, lo que refleja una capacidad predictiva muy alta. Además, la validación cruzada confirmó la estabilidad del modelo, con un promedio de accuracy de 98.43% y una desviación estándar baja ($\pm 1.87\%$).

Figura 3. Importancia de las variables en el modelo



Fuente: Elaboración propia en Google Colab con datos de Kaggle (Students' Social Media Addiction, Adil Shamim, 2025).

Los resultados permiten identificar patrones comunes en los estudiantes clasificados como adictos al uso de redes sociales. Las variables con mayor importancia en el modelo fueron el Mental_Health_Score, Conflicts_Over_Social_Media, Affects_Academic_Performance y Sleep_Hours_Per_Night, lo cual revela que los factores psicológicos y académicos tienen un peso significativo en la predicción.

Figura 4. Regla de adicción

Regla N°	Condiciones (Resumen)	Predicción
1	Affects > 0.50, Sleep ≤ 7.15, Age ≤ 20.50, Platform ≤ 3.50, MH ≤ 6.50, MH ≤ 5.50	Adicto
2	Affects > 0.50, Sleep ≤ 7.15, Age ≤ 20.50, Platform ≤ 3.50, MH ≤ 6.50, MH > 5.50, Country ≤ 80.50	Adicto
3	Affects > 0.50, Sleep ≤ 7.15, Age ≤ 20.50, Platform ≤ 3.50, MH ≤ 6.50, MH > 5.50, Country > 80.50, Relationship ≤ 1.50, Country > 95.50	Adicto
4	Affects > 0.50, Sleep ≤ 7.15, Age ≤ 20.50, Platform ≤ 3.50, MH ≤ 6.50, MH > 5.50, Country > 80.50, Relationship > 1.50	Adicto

5	Affects > 0.50, Sleep \leq 7.15, Age \leq 20.50, Platform > 3.50	Adicto
6	Affects > 0.50, Sleep \leq 7.15, Age > 20.50	Adicto
7	Affects > 0.50, Sleep > 7.15, MH \leq 6.50, Sleep \leq 7.65, Platform \leq 1.50	Adicto
8	Affects > 0.50, Sleep > 7.15, MH \leq 6.50, Sleep \leq 7.65, Platform > 1.50, Academic \leq 1.00	Adicto
9	Affects > 0.50, Sleep > 7.15, MH \leq 6.50, Sleep \leq 7.65, Platform > 1.50, Academic > 1.00, Sleep \leq 7.50, Country > 89.50	Adicto
10	Affects > 0.50, Sleep > 7.15, MH \leq 6.50, Sleep \leq 7.65, Platform > 1.50, Academic > 1.00, Sleep > 7.50, Country > 89.50	Adicto
11	Affects > 0.50, Sleep > 7.15, MH \leq 6.50, Sleep > 7.65	Adicto

Las reglas generadas por el modelo revelan combinaciones de condiciones que conducen a una clasificación de adicción. Por ejemplo, estudiantes que reportan que el uso de redes sociales afecta su rendimiento académico, duermen menos de 7.15 horas por noche, tienen un puntaje de salud mental bajo, y están en edades tempranas (\leq 20 años), presentan una alta probabilidad de ser clasificados como adictos. Esto sugiere una relación directa entre la adicción, el bajo bienestar emocional y el rendimiento académico.

Estos hallazgos pueden ayudar a las instituciones educativas y a los responsables de políticas estudiantiles a diseñar estrategias de intervención temprana. Programas de apoyo psicológico, promoción de buenos hábitos de sueño y campañas para concientizar sobre el uso responsable de redes sociales podrían ser implementados para reducir los efectos negativos detectados y apoyar a los estudiantes en riesgo.

7. CONCLUSIONES Y PRÓXIMOS PASOS

El análisis realizado permitió identificar patrones consistentes en los estudiantes clasificados como adictos a las redes sociales. Uno de los principales hallazgos es que la percepción de que las redes afectan negativamente el rendimiento académico, combinada con bajos niveles de sueño (≤ 7.15 horas por noche), es un fuerte predictor de adicción. Asimismo, variables como la edad, el puntaje de salud mental, y la plataforma más utilizada también juegan un papel relevante en la clasificación.

Los resultados revelan que los estudiantes más jóvenes (≤ 20.5 años), con puntajes bajos de salud mental y menor tiempo de descanso, tienen una mayor probabilidad de ser clasificados como adictos. Además, factores contextuales como el país de residencia, el nivel académico y el estado de relación también influyen en ciertos escenarios. Estas reglas pueden ser útiles para diseñar estrategias de intervención dirigidas a grupos de riesgo específicos.

Se recomienda ampliar el análisis utilizando una muestra más diversa y representativa de estudiantes de distintas regiones y niveles educativos, con el fin de mejorar la generalización del modelo. También sería útil explorar técnicas más avanzadas como XGBoost para comparar su rendimiento con los modelos actuales y evaluar si se mejora la precisión y robustez de las predicciones.

Como próximos pasos, se sugiere incorporar variables psicológicas y socioeconómicas adicionales (como ansiedad, autoestima, tipo de uso de redes, etc.), que podrían enriquecer el análisis y proporcionar una visión más integral del fenómeno. Finalmente, implementar visualizaciones interactivas y un panel de control (dashboard) permitiría comunicar estos resultados de forma más efectiva a instituciones educativas y responsables de políticas públicas.

8. REFERENCIAS

Shamim, A. (2025). Social Media Addiction vs Relationships. Kaggle.
Recuperado de

<https://www.kaggle.com/datasets/adilshamim8/social-media-addiction-vs-relationships>