

# ¿Qué hace que un estudiante rinda mejor? Un análisis basado en datos reales



Ángel Velázquez Bolívar

29/05/2025

# Índice

1	Introducción . . . . .	2
2	Descripción de los datos . . . . .	2
3	Exploración inicial: ¿qué dicen los datos? . . . . .	3
4	¿Qué factores influyen en el rendimiento académico? . . . . .	5
5	Segmentación de estudiantes . . . . .	10
6	Conclusiones y recomendaciones finales . . . . .	13

# 1 Introducción

Este informe presenta los resultados de un estudio cuyo objetivo ha sido conocer mejor a los estudiantes: ¿cómo influyen sus hábitos diarios en su rendimiento académico? ¿Podemos predecir si un estudiante va bien o mal en sus estudios en función de su estilo de vida?

Para responder a estas preguntas, se han analizado datos recopilados de estudiantes, incluyendo información sobre sus horas de estudio, uso de redes sociales, sueño, salud mental, entre otros factores. A partir de estos datos, se han aplicado técnicas estadísticas y análisis de datos para extraer conclusiones útiles.

Todo el análisis se ha llevado a cabo utilizando el lenguaje de programación R, una herramienta ampliamente reconocida en el ámbito científico y académico por su potencia para el tratamiento y visualización de datos.

Aunque el análisis se ha realizado con herramientas técnicas, este informe está diseñado específicamente para el cliente como lector no especializado. Aquí encontrarás explicaciones sencillas, ejemplos visuales y las conclusiones más relevantes, sin necesidad de tener conocimientos matemáticos o estadísticos.

Los resultados obtenidos pueden ser de gran utilidad para mejorar el rendimiento de los estudiantes, detectar situaciones de riesgo o incluso adaptar estrategias educativas a distintos perfiles.

A lo largo del informe iremos explorando qué hemos descubierto y cómo se puede aprovechar esta información.

## 2 Descripción de los datos

El estudio se ha basado en la información recogida de una muestra de estudiantes. A cada estudiante se le pidió información sobre una serie de aspectos de su vida diaria. En total, se han analizado las siguientes variables:

Variable	Descripción
Edad	Edad del estudiante
Sexo	Sexo del estudiante (Hombre/Mujer/Otro)
Horas_estudio_diario	Horas de estudio diario
Horas_redes_sociales	Horas de uso de redes sociales diario
Horas_netflix	Horas de uso de Netflix diario
Trabajo	Indica si el estudiante trabaja o no
Asistencia	Porcentaje de asistencia a clase
Horas_sueño	Horas diarias de sueño
Dieta	Calidad de la alimentación
Dias_ejercicio_semanal	Días de ejercicio a la semana
Educacion_padres	Nivel educativo de los padres
Calidad_internet	Calidad del internet del estudiante
Salud_mental	Indica la salud mental del estudiante
Actividades_extracurriculares	Realización de actividades extracurriculares
Calificacion	Nota media del estudiante

Estas variables nos permiten tener una visión amplia del estilo de vida de cada estudiante, desde aspectos académicos (como el estudio y la asistencia), hasta factores personales como el descanso, el ocio, la salud mental y la calidad de conexión a internet.

En las siguientes secciones, veremos cómo se ha utilizado esta información para detectar patrones y generar modelos que nos permitan predecir el rendimiento académico de los estudiantes.

### 3 Exploración inicial: ¿qué dicen los datos?

Antes de aplicar modelos más complejos, realizamos una exploración general de los datos para entender mejor cómo se comportan las variables y cómo se relacionan entre ellas.

#### ¿Cómo es la distribución de calificaciones?

Una de las primeras preguntas que nos hicimos fue: *¿qué tan bien están rindiendo los estudiantes en general?* En la Figura 1 se muestra cómo se distribuyen las calificaciones entre todos los estudiantes encuestados.

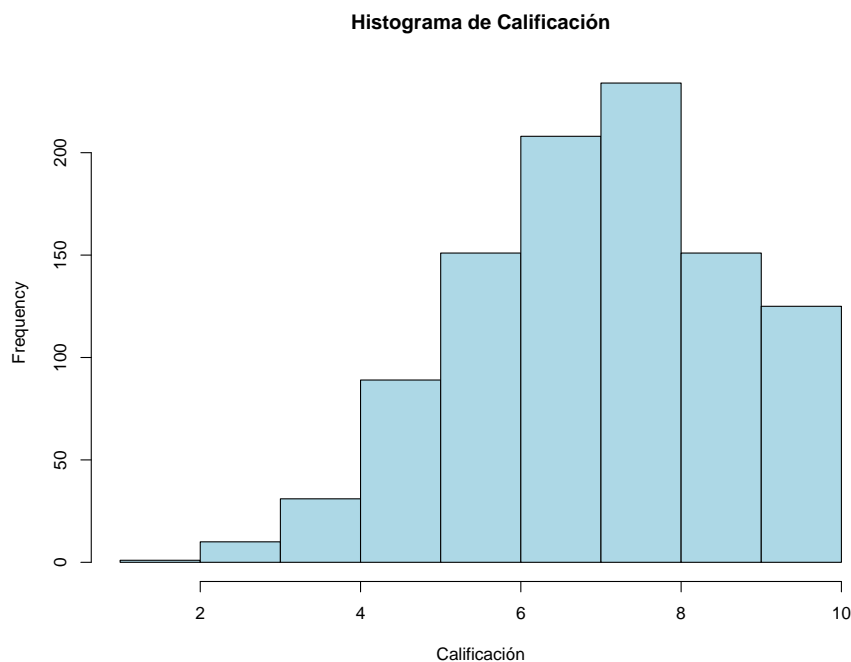


Figura 1: Distribución de las calificaciones

Como puede observarse, la mayoría de estudiantes se sitúan entre el 5 y el 8 de nota media, lo que indica un rendimiento general moderado. Sin embargo, también hay estudiantes con notas muy altas (por encima de 9) y otros, aunque menos, por debajo del aprobado.

Se realizó también un estudio descriptivo de los datos con el fin de conocer mejor los hábitos y características de los estudiantes. Se encontró que la edad media de los estudiantes es de 20.5 años y que en promedio dedican 3.55 horas diarias al estudio. La calificación media es de 6.96 puntos.

## Relaciones entre variables

Para analizar posibles relaciones entre los hábitos de vida y las calificaciones, calculamos las correlaciones entre las variables numéricas. El siguiente gráfico (Figura 2) muestra cómo se relacionan entre sí algunas variables clave.

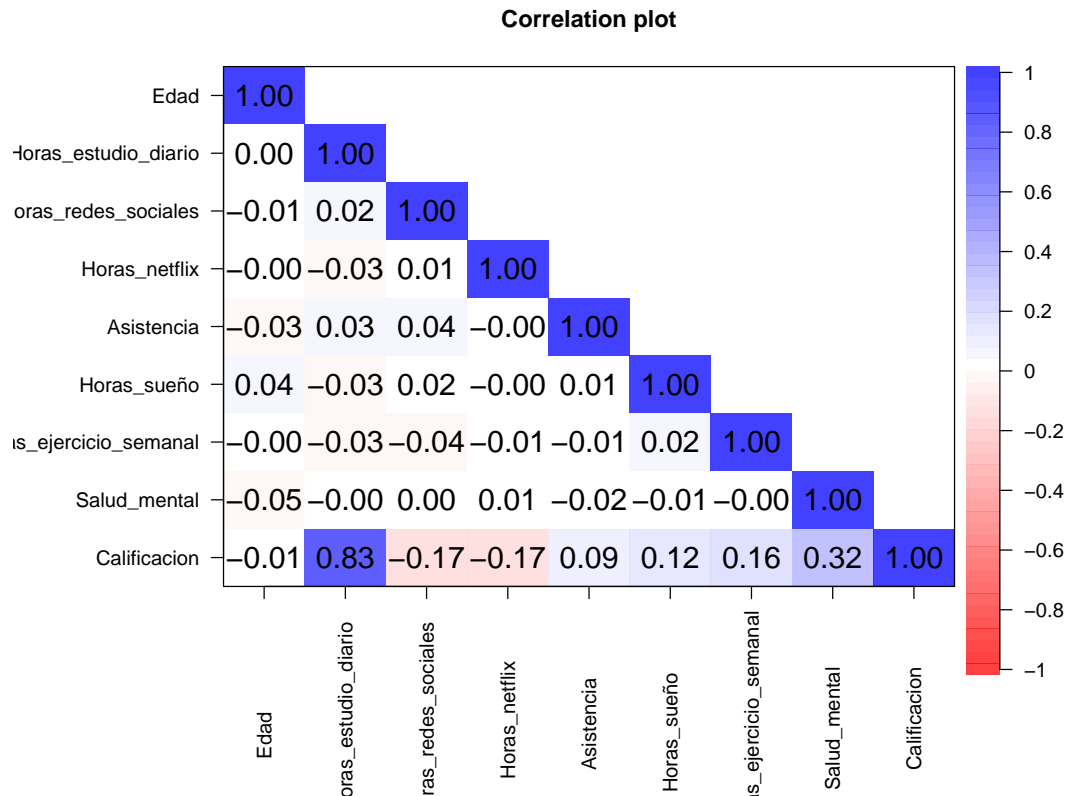


Figura 2: Relaciones entre las variables

En este gráfico, los colores más oscuros indican una relación más fuerte entre dos variables. Por ejemplo, se observa una relación positiva entre las **Horas\_estudio\_diario** y la **Calificación**, lo que sugiere que estudiar más está asociado a sacar mejores notas. También se aprecia que un mayor uso de redes sociales o plataformas como Netflix tiende a estar asociado con calificaciones algo más bajas, aunque estas relaciones son más débiles.

El resto variables apenas muestran relaciones entre sí, lo que sugiere que tendrán una menor importancia en nuestros modelos.

Estos primeros análisis nos ayudaron a identificar qué aspectos del estilo de vida parecen tener más influencia en el rendimiento académico, y cuáles podrían ser útiles en los modelos predictivos que se explican en las siguientes secciones.

## 4 ¿Qué factores influyen en el rendimiento académico?

Con el fin de entender qué características de los estudiantes afectan más a sus calificaciones, se construyeron diversos modelos explicativos y predictivos. Para facilitar el análisis, se utilizó tanto la nota numérica original (de 0 a 10), como una versión categorizada en tres niveles de rendimiento:

- **Bajo:** calificaciones entre 0 y 5.

- **Medio:** calificaciones entre 5 y 8.
- **Alto:** calificaciones entre 8 y 10.

Esta clasificación permitió abordar el problema desde dos enfoques diferentes:

1. Explicar qué variables influyen en la nota del estudiante (regresión).
2. Predecir si un estudiante pertenece al grupo de bajo, medio o alto rendimiento (clasificación).

## Modelos de regresión: ¿Qué influye en la nota?

Se probaron distintos modelos que intentaban explicar la nota media de los estudiantes a partir de todas las variables recogidas. El primer modelo incluyó todas las variables disponibles, pero tras aplicar métodos de mejora y simplificación, se seleccionó un modelo más reducido con solo 9 variables que ofrecía resultados prácticamente igual de buenos.

### Variables incluidas en el modelo final:

- Horas\_estudio\_diario
- Horas\_redes\_sociales
- Horas\_netflix
- Asistencia
- Horas\_sueño
- Dias\_ejercicio\_semanal
- Calidad\_internet
- Salud\_mental
- Actividades\_extracurriculares

Este modelo final logró explicar cerca del 90% de la variabilidad en las notas, lo que indica que las variables seleccionadas capturan muy bien las diferencias de rendimiento entre estudiantes.

Al aplicar este modelo sobre datos nuevos (no usados para entrenarlo), obtuvo una correlación de aproximadamente 0.94 entre las notas reales y las predichas. Una correlación cercana a 1 indica una fuerte relación, por lo que esto implica una muy buena capacidad predictiva. En la Figura 3 se pueden ver las predicciones realizadas por este modelo frente a las calificaciones reales. En general, se observa una fuerte alineación de los puntos en torno a la diagonal, lo cual indica que las predicciones realizadas por el modelo son correctas.

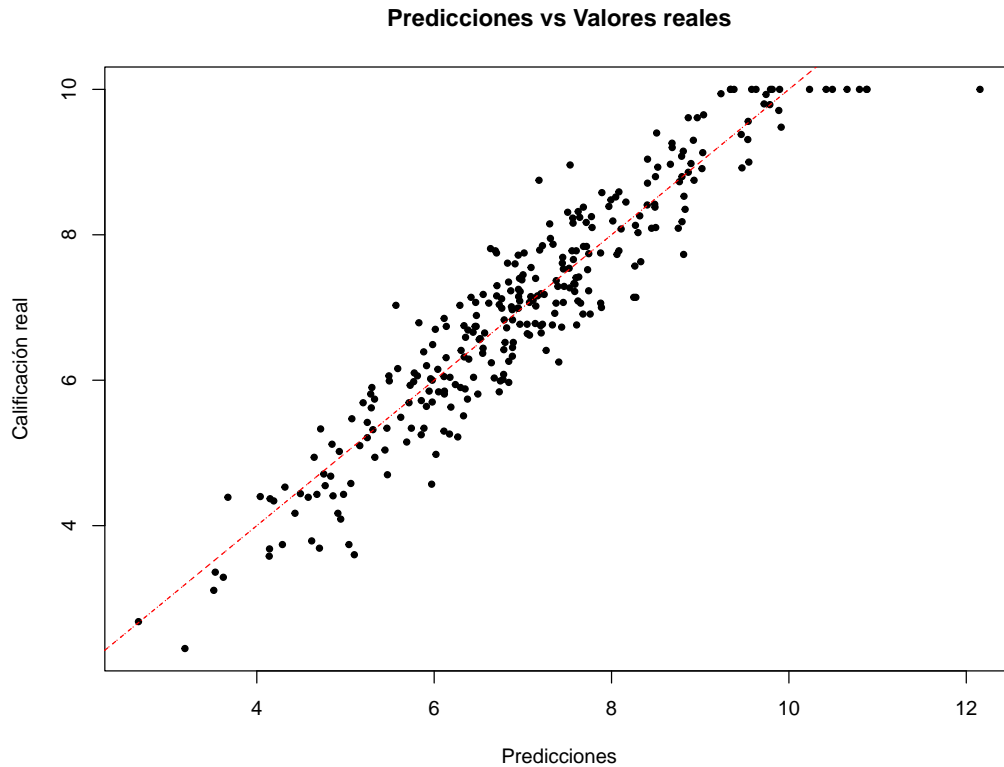


Figura 3: Relaciones entre las variables

### Conclusiones clave:

- Los estudiantes que dedican más tiempo al estudio y tienen una buena asistencia a clase tienden a obtener mejores calificaciones.
- La salud mental también juega un papel importante: quienes se sienten mentalmente bien suelen rendir mejor.
- En cambio, un mayor consumo de redes sociales y Netflix se asocia con notas más bajas.

### Modelos de clasificación: ¿Podemos predecir si un estudiante será exitoso?

Además de predecir la nota exacta, se desarrollaron modelos para anticipar si un estudiante se encuentra en el grupo de bajo, medio o alto rendimiento, utilizando sus características personales.

### Árbol de decisión

Este método genera una serie de reglas simples, similares a un esquema de “sí/no”, que permiten clasificar a los estudiantes. El árbol generado (Figura 4) permite visualizar qué combinación de variables lleva a un estudiante a una categoría u otra.



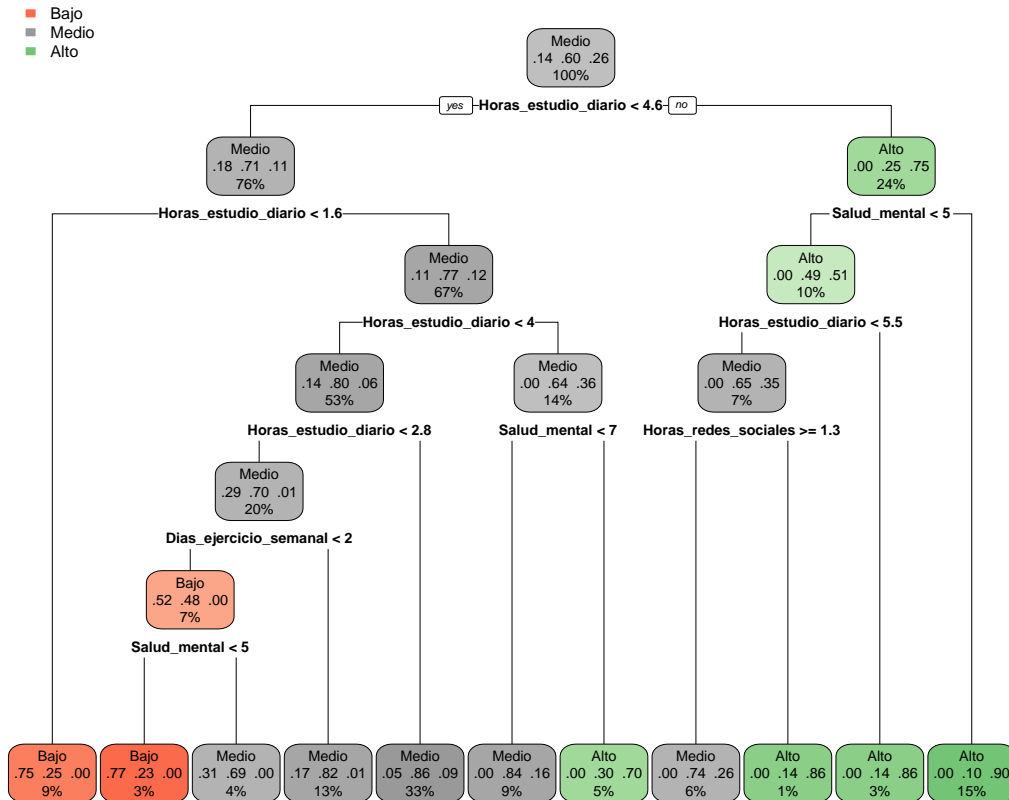


Figura 4: Árbol de decisión para predecir el rendimiento académico

Este modelo logró una precisión del 79.7% en nuevos datos. Aunque sencillo, permitió detectar patrones como:

- Los estudiantes con pocas horas de estudio (menos de 1.6 horas diarias) y baja salud mental tienden a clasificarse como de *bajo rendimiento*.
- En la mayoría de nodos intermedios predomina la clase *Medio*, lo que indica que es el perfil más común.
- Los estudiantes con un número elevado de horas de estudio (más de 5.5 horas diarias), combinadas con bajo uso de redes sociales o altos niveles de salud mental, se asocian a una mayor probabilidad de pertenecer a la clase *Alto*.

## Random Forest

Para mejorar la precisión, se utilizó una técnica más avanzada: **random forest**, que combina muchos árboles simples para tomar decisiones más robustas. Este modelo alcanzó una precisión del 81% y confirmó la importancia de las siguientes variables:

- Horas\_estudio\_diario
- Salud\_mental
- Horas\_redes\_sociales

Tal y como muestra la Figura 5, en la que se representa la importancia de cada variable tuvo en las predicciones.

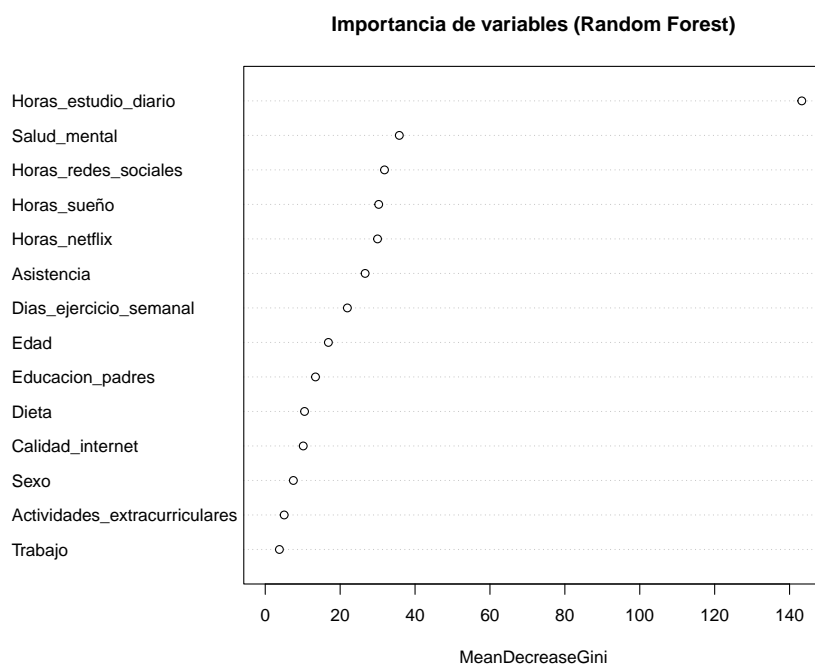


Figura 5: Importancia de las variables en el modelo Random Forest

## Boosting

Por último, se probó una técnica llamada **boosting**, que combina árboles entrenados de manera secuencial, aprendiendo de los errores anteriores. Este modelo fue el que obtuvo mejores resultados, alcanzando una precisión del 81.3%, ligeramente superior a random forest. En la Figura 6 se muestra la importancia de cada variable en este modelo, de igual forma que se mostró para el modelo random forest.

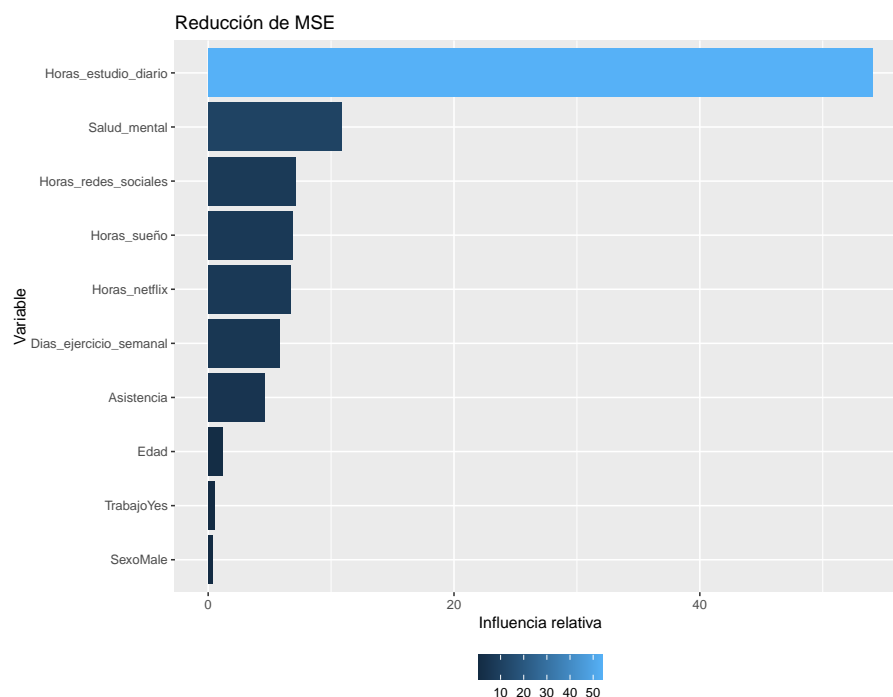


Figura 6: Importancia de las variables en el modelo de Boosting

Los resultados refuerzan la conclusión de que las variables más importantes para predecir el éxito académico son las horas de estudio, la salud mental y el tiempo pasado en redes sociales.

En resumen, tanto los modelos de regresión como los de clasificación coinciden en que ciertos hábitos y factores personales están fuertemente asociados con el rendimiento académico. Esta información puede ser útil para orientar recomendaciones personalizadas o desarrollar programas de mejora del rendimiento.

## 5 Segmentación de estudiantes

Además de predecir calificaciones, también se exploró si era posible agrupar a los estudiantes en perfiles similares según sus hábitos y rutinas, sin utilizar su nota como referencia. A este tipo de análisis se le llama clasificación no supervisada, ya que los grupos no están definidos de antemano.

El objetivo es encontrar patrones naturales en los datos que nos permitan detectar tipos de estudiantes con características similares entre sí y distintas respecto a otros grupos.

### ¿Qué variables se usaron?

Para segmentar a los estudiantes, se seleccionaron las siguientes variables numéricas directamente relacionadas con sus hábitos cotidianos:

- Horas\_estudio\_diario

- Horas\_redes\_sociales
- Horas\_netflix
- Horas\_sueño

Estas variables ofrecen una visión general del uso del tiempo por parte de cada estudiante.

## ¿Cómo se agruparon los estudiantes?

Se utilizaron dos métodos diferentes para formar grupos:

- **Clustering jerárquico (Ward):** Agrupa a los estudiantes de forma progresiva, combinando los más parecidos entre sí.
- **K-medias:** Parte de un número de grupos fijado y asigna a cada estudiante el grupo más parecido a su perfil.

Se optó por formar 4 grupos, un número que ofrecía un equilibrio razonable entre simplicidad y diferenciación.

## ¿Qué grupos se detectaron?

El algoritmo de k-medias fue el que obtuvo mejores resultados. Aunque los grupos formados no eran completamente separados entre sí, sí permitió identificar ciertos patrones de comportamiento.

En la Figura 7 se muestra una representación visual de los grupos identificados. Cada punto representa a un estudiante, y los colores corresponden a los diferentes grupos. Aunque hay solapamiento, se aprecian al menos dos grupos claramente diferenciados.

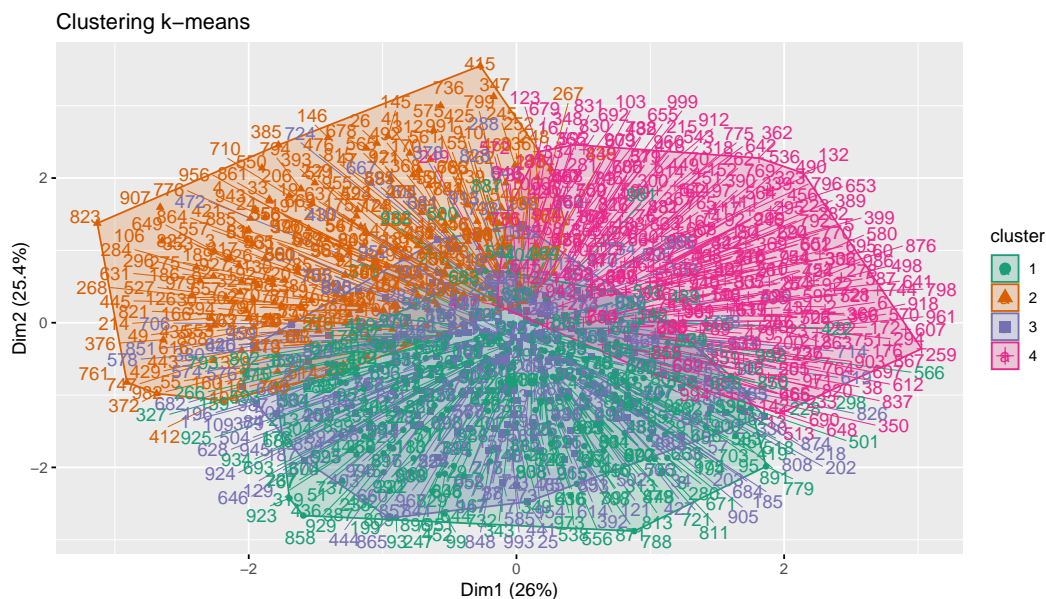


Figura 7: Agrupación de estudiantes según sus hábitos (modelo k-medias)

## Perfil medio de cada grupo

Para interpretar los grupos formados, se calcularon las medias (tipificadas) de cada variable dentro de cada grupo. A partir de ello, se pueden describir los perfiles siguientes:

- Grupo 1: Presenta valores bajos en horas de sueño y en uso de redes sociales, pero un valor elevado en horas de Netflix. Su dedicación al estudio es ligeramente inferior a la media. Este grupo podría representar a estudiantes que ven series hasta altas horas de la noche.
- Grupo 2: Se caracteriza por tener valores muy altos en horas de redes sociales, Netflix y sueño, mientras que el tiempo de estudio es claramente inferior a la media. Este grupo podría representar a estudiantes con un estilo de vida poco enfocado en el estudio y que dedican mucho tiempo a las pantallas.
- Grupo 3: Muestra los valores más bajos en casi todas las variables, especialmente en uso de redes sociales y Netflix, así como también por debajo de la media en estudio y sueño. Podría tratarse de un grupo con baja actividad general.
- Grupo 4: Se diferencia por dedicar mucho más tiempo al estudio que el resto, con valores elevados también en uso de redes sociales, pero con menos horas de sueño y bajo consumo de Netflix. Este grupo parece representar a estudiantes más enfocados académicamente, aunque con menor descanso.

## ¿Qué tan fiables son estos grupos?

Para comprobar la calidad del agrupamiento, se usó una medida conocida como coeficiente de silueta. Esta métrica mide la similitud de un objeto con los demás objetos de su grupo en comparación con otros grupos. Su valor va de -1 (mala agrupación) a 1 (muy buena agrupación). En este caso, se obtuvo un valor de 0.17, lo que indica que la estructura encontrada es *débil*, aunque algo mejor que una agrupación completamente aleatoria.

De esta forma, se lograron identificar algunos patrones generales en los hábitos de los estudiantes, pero no se encontraron grupos bien definidos ni claramente diferenciados. Esto sugiere que los estilos de vida de los estudiantes son diversos y que no es fácil encasillar a los estudiantes en perfiles fijos únicamente a partir de sus rutinas diarias. No obstante, este análisis fue útil para explorar tendencias generales y entender mejor las diferencias entre perfiles de estudiantes.

## 6 Conclusiones y recomendaciones finales

Este estudio nos ha permitido comprender mejor cómo influyen los hábitos y condiciones personales en el rendimiento académico de los estudiantes.

### Principales conclusiones

- **El tiempo de estudio es clave.** Los estudiantes que dedican más horas al estudio diario tienden, de forma clara y consistente, a obtener mejores calificaciones. Esta fue la variable más influyente en todos los modelos analizados.
- **La salud mental marca la diferencia.** Aquellos estudiantes que reportan una buena salud mental también presentan mejores resultados académicos, lo que destaca la importancia del bienestar emocional para el aprendizaje.
- **El uso excesivo de redes sociales y plataformas de entretenimiento puede afectar negativamente.** Aunque estas actividades no son perjudiciales por sí mismas, un uso elevado tiende a asociarse con peores resultados académicos.
- **Es posible predecir el rendimiento de los estudiantes con buena precisión.** Los modelos desarrollados alcanzaron hasta un 81 % de acierto, lo que demuestra que, con la información adecuada, se pueden anticipar los resultados académicos y actuar a tiempo.
- **Los perfiles de estudiantes no son fácilmente clasificables.** Aunque se identificaron algunos grupos con características comunes, la diversidad de hábitos entre los alumnos hace que la segmentación en grupos fijos no sea totalmente precisa.

### ¿Qué implicaciones tiene esto?

Los resultados obtenidos ofrecen una base sólida para diseñar estrategias de mejora educativa:

- **Recomendaciones personalizadas:** Con herramientas como las desarrolladas en este estudio, se pueden diseñar sistemas de apoyo que identifiquen estudiantes con riesgo de bajo rendimiento y ofrecerles orientación personalizada.
- **Promoción de hábitos saludables:** Fomentar el estudio regular, cuidar la salud mental y gestionar mejor el tiempo dedicado al ocio digital pueden ser claves para mejorar los resultados académicos.

### Reflexión final

Este informe demuestra que detrás de los números y las notas hay patrones que pueden ser comprendidos, analizados y, sobre todo, utilizados para mejorar. La combinación de datos y modelos estadísticos nos permite pasar de la intuición al conocimiento real.

Este tipo de análisis ofrece un gran potencial para instituciones educativas y profesores, ya que permite tomar decisiones basadas en evidencia real.