

Informe Final - Diplomado en Ciencia de Datos

1. Título del Proyecto

¿Cómo predecir el promedio de un estudiante según sus hábitos de estudio y motivación?

2. Objetivo

El objetivo de este proyecto fue crear un modelo que ayudara a predecir el promedio de calificaciones de los estudiantes, usando datos como sus horas de estudio, asistencia, motivación, uso de la plataforma, entre otros. Para este trabajo usé un dataset con datos aleatorios, inventados especialmente para el proyecto, ya que no logré acceder a un conjunto de datos reales, y el que me proporcionó mi empresa (Filadd Chile) es muy denso y difícil de manejar para mi nivel de experiencia en el área. La idea era ver si es posible anticipar el rendimiento con esa información.

3. Introducción

Contexto

Hoy en día, las instituciones educativas tienen muchos datos sobre sus estudiantes. Con esta información, se pueden crear modelos que ayuden a entender cómo y por qué algunos estudiantes rinden mejor que otros.

Qué buscaba el modelo

El modelo buscaba predecir el promedio de un estudiante. Es decir, al ingresar los datos de un estudiante nuevo, el sistema intenta decir cuánto sería su promedio final.

4. Descripción de los Datos

De dónde salieron los datos

Los datos utilizados fueron creados específicamente para este trabajo, con el objetivo de simular un grupo realista de estudiantes. Es importante recalcar que al tratarse de datos inventados, esto influye directamente en la precisión del modelo y sus predicciones.

Me interesa desarrollar un segundo proyecto en el futuro, esta vez con un conjunto de datos reales. ¿Es posible lograrlo? Creo que sí. Tengo un emprendimiento educativo que me mantiene en contacto con al menos 10 estudiantes, algunos muy motivados y otros no tanto.

Podría aplicar una encuesta para recolectar información antes y después de la PAES, incluyendo variables relevantes para el rendimiento.

Este tipo de levantamiento de datos me permitiría construir un modelo más cercano a la realidad, con resultados más útiles y aplicables. Sería un proyecto interesante y valioso de realizar como una segunda etapa.

Qué tenían los datos

Los datos incluían información como la edad, cuántas horas estudian, cuánto asisten a clases, si participan en actividades, qué tan motivados están, si tienen acceso a recursos y cuánto usan la plataforma. También incluían el promedio de notas y una categoría de rendimiento general (1, 2 o 3).

Qué hice con los datos

Se eliminaron algunas columnas que no se usaron, como el ID. Se dejó solo lo que servía para predecir el promedio. No había datos faltantes, así que no fue necesario limpiarlos mucho.

5. Modelo que usé

Qué modelo utilicé

Usé un modelo simple llamado “regresión lineal”. Sirve para predecir un número (en este caso, el promedio) a partir de otros valores.

Cómo lo preparé

El modelo usó la mayoría de los datos (80%) para aprender y el resto (20%) para probar si había aprendido bien. No se ajustaron muchas cosas, porque este modelo es simple y no tiene muchas configuraciones.

6. Cómo le fue al modelo

Para saber si el modelo funciona bien, se compararon los promedios reales con los promedios que el modelo predijo. Se usaron algunas medidas que ayudan a saber si el error es grande o pequeño. Los resultados no fueron perfectos, pero tenían lógica: los estudiantes con más estudio, motivación y asistencia solían tener promedios más altos.

7. Interpretación de los resultados

Probé el modelo con tres estudiantes inventados (insisto, creo que funcionaría mucho mejor con una aplicación real. Es algo que volveré a implementar):

- Uno tenía buena motivación y asistencia: su promedio predicho fue 11.88
- Otro tenía baja motivación y pocas horas de estudio: su promedio fue 9.96
- Uno muy motivado y que estudiaba mucho: su promedio fue 12.86

Aunque los resultados tienen sentido, los **promedios salieron muy altos (más de 10)**, lo cual no es posible en la vida real (para este trabajo utilicé una escala de 1 a 10). Esto pasó porque los datos eran inventados/aleatorios y no estaban ajustados a una escala real.

8. Conclusiones

A pesar de que los datos eran ficticios, este proyecto me ayudó a entender cómo funcionan los modelos de predicción. Aprendí a preparar los datos, a entrenar un modelo y a probarlo con ejemplos. Si en el futuro tuviera datos reales, podría usar lo aprendido para hacer predicciones más cercanas a la realidad y tomar mejores decisiones en el ámbito educativo.