



Bootcamp de Data Science

Bootcamp Data Science 1

1



Ejercicio semana 8

Miriam Ramírez Muñoz

Bootcamp Data Science 2

2



Situación:

Eres un científico de datos en una empresa de comercio electrónico. La empresa ha recolectado una gran cantidad de datos sobre las interacciones de los usuarios con el sitio web durante los últimos años, generando terabytes de datos. Tu tarea es analizar estos datos para identificar patrones de compra y, en base a ello, mejorar las recomendaciones de productos a los usuarios.

1. Big Data: Dada la gran cantidad de datos recolectados, ¿cuál considerarías una herramienta adecuada para almacenar y procesar estos datos: una base de datos relacional tradicional o una plataforma como Hadoop o Spark? Justifica tu elección.

3



2. Reducción de la Dimensionalidad: Una vez que tienes los datos listos para el análisis, te das cuenta de que hay cientos de características para cada usuario (edad, género, historial de navegación, historial de compras, etc.). ¿Qué técnica usarías para reducir la dimensionalidad de estos datos y por qué?

3. Modelado: Quieres construir un modelo que prediga si un usuario compraría un producto basado en sus características. ¿Qué tipo de modelo usarías: un modelo de regresión o de clasificación? Justifica tu elección.

4. Ajuste del Modelo: Una vez que hayas decidido el tipo de modelo, ¿qué técnica utilizarías para asegurarte de que tu modelo está bien ajustado y no está sobreajustado o subajustado?

5. Evaluación del Modelo: Finalmente, antes de implementar tu modelo en producción, necesitas evaluar su rendimiento. ¿Qué métricas usarías para evaluar un modelo de clasificación y por qué?

4