

## PARCIAL III CORTE

### 1. ¿Qué es un libro electrónico IPYNB?

- Los libros electrónicos IPYNB (también conocidos como Jupyter Notebooks) son un formato de archivo interactivo que le permite crear y compartir documentos que contienen código, texto explicativo, gráficos y otros elementos multimedia. Consta de celdas, cada una de las cuales puede contener código ejecutable o texto con formato Markdown, y permite formatear texto, imágenes y enlaces. Estos libros proporcionan una forma interactiva de explorar y analizar datos, lo que le permite ejecutar y modificar código en tiempo real y ver los resultados al instante.

### 2. Describa brevemente la funcionalidad de las siguientes librerías: (i) Pandas ; (ii) NumPy ; (iii) Matplotlib

- **Biblioteca Pandas:** Pandas es una biblioteca de Python que se utiliza para el análisis y la manipulación de datos. Proporciona estructuras de datos flexibles (Series y DataFrames) para ordenar y manejar datos. Le permite leer y escribir datos en diferentes formatos, limpiar y modificar datos, escoger y filtrar datos y ejecutar operaciones de agregación y resumen. Pandas se integra con otras bibliotecas populares de Python, lo que lo hace poderoso y versátil en proyectos de ciencia de datos.
- **Biblioteca NumPy:** NumPy es una biblioteca de Python para procesamiento numérico y científico. Proporciona un objeto llamado "ndarray" para trabajar con matrices multidimensionales y realizar operaciones matemáticas eficientes en ellas. NumPy proporciona funciones matemáticas, indexación avanzada, transmisión e integración con otras bibliotecas científicas. Es la base de la computación científica en Python.
- **Biblioteca Matplotlib:** Matplotlib es una biblioteca de visualización de datos de Python que le permite crear gráficos personalizados de alta calidad. Puede crear gráficos básicos (líneas, barras, dispersión) y especializados (contorno, violín, cuadro), personalizar su aspecto y generar gráficos 3D. También proporciona interactividad y animaciones para investigar datos de manera dinámica. Matplotlib facilita la representación de datos de forma eficiente y atractiva en Python.

### 3. ¿Cuáles son los principales algoritmos de recomendación utilizados en las aplicaciones web? Explique brevemente

- En las aplicaciones web, se utilizan varios algoritmos de recomendación:
  - **Filtrado híbrido:** este enfoque combina varios algoritmos de recomendación para obtener resultados más precisos y personalizados.

Puede combinar el filtrado colaborativo y el filtrado basado en contenido para aprovechar ambos enfoques.

- **Filtrado basado en contenido:** este algoritmo utiliza características o atributos de elementos para hacer recomendaciones. Examine las características de los artículos previamente consumidos o calificados por los usuarios y encuentre artículos similares basados en estas características.
- **Filtros colaborativos:** este algoritmo utiliza información de las preferencias del usuario para hacer recomendaciones. Se basa en la idea de que si a dos usuarios les gustaron elementos similares en el pasado, es probable que tengan gustos similares en el futuro.

4. ¿Cuáles son las bases de datos utilizadas mayoritariamente en los procesos de análisis de datos que utilizan técnicas de machine learning?

- Algunas de las bases de datos utilizadas mayoritariamente en este contexto son:
  - **PostgreSQL:** PostgreSQL es otra base de datos relacional de código abierto que se utiliza comúnmente en proyectos de análisis de datos. Ofrece un amplio rango de características y soporta consultas complejas, lo que lo hace adecuado para aplicaciones de machine learning que requieren un alto grado de flexibilidad en las consultas.
  - **MySQL:** MySQL es una base de datos relacional de código abierto ampliamente utilizada. Es conocido por su rendimiento, escalabilidad y facilidad de uso. Se utiliza en muchos proyectos de aprendizaje automático para almacenar y administrar grandes conjuntos de datos estructurados.
  - **Apache Cassandra:** Cassandra es una base de datos distribuida altamente escalable y tolerante a fallas. Es particularmente adecuado para aplicaciones de big data y aprendizaje automático donde la escalabilidad y el rendimiento son críticos. Cassandra se usa a menudo para almacenar grandes conjuntos de datos que requieren alta disponibilidad y rendimiento en tiempo real