

Prueba técnica para Ingeniero de Machine Learning

Objetivo

Evaluar tus habilidades técnicas en el desarrollo de sistemas de recomendación, entrenamiento de redes neuronales, y exposición de servicios mediante APIs. Esta prueba busca comprender tu enfoque, capacidad de justificación de decisiones, y habilidad para implementar soluciones completas.

Deberás crear un sistema de recomendación que utilice al menos una red neuronal. No es obligatorio que el sistema sea estrictamente basado en redes neuronales; puedes emplear un enfoque híbrido siempre y cuando uses al menos una red neuronal como parte de la solución. El sistema debe recomendar productos (identificados por la columna `parent_product_id`) a los clientes (identificados por la columna `customer_id`), utilizando los datos descritos en la sección "Datos para la prueba".

Entregables

1. Notebook de Python en formato ipynb:

- Debe contener el desarrollo y validación del modelo de red neuronal y el sistema de recomendación.
- Documentar de manera clara los pasos seguidos, las decisiones tomadas y la metodología empleada.

2. API en Python:

- Implementar una API que reciba un `customer_id` y devuelva al menos 2 recomendaciones personalizadas.
- Cada recomendación debe incluir al menos los campos `producto_id`, `producto_title`
- Dockerizar la API para que pueda ejecutarse en cualquier entorno.
- Incluir un archivo `README.md` con las instrucciones necesarias para construir y ejecutar el contenedor Docker.

3. Enlace de entrega:

- Subir todos los entregables a un servicio de almacenamiento (Google Drive o One Drive)
- Compartir el enlace a los correos aortizp@glingeriasas.com y dpineda@magnetoglobal.com.

- Fecha límite de entrega: **viernes 14 de marzo de 2025 a las 23:59:59 (hora de Bogotá).**

Criterios de evaluación

Se evaluará la completitud de la entrega, la validez técnica de los procedimientos aplicados y su justificación.

Una vez entregados los resultados, se programará una reunión en la que deberás explicar y justificar tu entrega. Durante esta sesión, evaluaremos tu capacidad para razonar sobre las decisiones tomadas y responder preguntas relacionadas con tu implementación.

Consideraciones Finales

- Eres libre de estructurar el notebook según consideres adecuado, siempre y cuando cuente con la documentación necesaria para entender el proceso realizado.
- No se espera que las métricas obtenidas sean perfectas, pero sí que expliques la elección del tipo de métricas, cómo las interpretaste y qué decisiones tomaste en base a ellas.
- Si consideras que hay información faltante, puedes realizar los supuestos necesarios para que puedas continuar con la prueba; sólo déjalos bien documentados en el notebook o en el archivo README.md.

Datos para la prueba

Los datos deben ser descargados del siguiente enlace: [fashion_reviews.zip](#).

Diccionario de Datos

A continuación, se presenta una breve descripción de las tablas disponibles en la base de datos `fashion_reviews.db`, junto con el diccionario de datos asociado a cada tabla. La base de datos fue creada usando `Sqlite v.3.45.3`, y cuenta con dos tablas: `reviews` y `products`. Las tablas están diseñadas para almacenar información relacionada con productos de moda y sus reseñas.

Tabla `reviews`

Descripción

La tabla `reviews` almacena las reseñas realizadas por los clientes sobre los productos. Cada registro incluye información sobre la calificación, el contenido de la reseña, las imágenes asociadas y datos del cliente que realizó la evaluación.

Diccionario de datos

Campo	Tipo	Descripción
review_score	float	Calificación de la reseña del producto (de 1.0 a 5.0).
title	str	Título de la reseña del usuario.
review_content	str	Cuerpo detallado de la reseña del usuario.
images	list	Enlaces de las imágenes asociadas a la reseña del producto.
product_id	str	ID único del producto reseñado.
parent_product_id	str	ID del grupo de productos relacionados (agrupa variaciones de color, estilo, etc).
customer_id	str	ID del cliente que realizó la reseña.
review_date	int	Fecha de la reseña en formato UNIX (timestamp).
upvotes	int	Número de votos útiles que recibió la reseña.
purchase_verified	bool	Indica si la compra fue verificada.

Tabla products

Descripción

La tabla products contiene información detallada sobre los productos de moda. Incluye datos como el nombre, la calificación promedio, el precio, las características, las imágenes, y las categorías a las que pertenece el producto.

Diccionario de datos

Campo	Tipo	Descripción
product_title	str	Nombre del producto.
average_rating	float	Calificación promedio del producto mostrada en la página del producto.
rating_count	int	Número de calificaciones recibidas por el producto.
features	list	Características del producto en formato de puntos clave.
description	list	Descripción detallada del producto.

price	float	Precio del producto en dólares estadounidenses (al momento del scraping).
images	list	Imágenes del producto con distintos tamaños (thumb, large, hi_res).
store_name	str	Nombre de la tienda que vende el producto.
categories	list	Categorías jerárquicas a las que pertenece el producto.
details	dict	Detalles del producto (materiales, marca, tallas, etc.).
parent_product_id	str	ID del grupo de productos relacionados (agrupa variaciones de color, estilo, etc).
bought_together	list	Productos recomendados para comprarse juntos desde la página del producto.

Estamos emocionados por conocer tu enfoque y cómo abor das este reto. ¡Muchos éxitos!