**Proyecto\_CineSugerencias**

**Proyecto Individual Nº1: Machine Learning Operations (MLOps)**

¡Bienvenidos a \*\*Proyecto\_CineSugerencias\*\*! En este primer proyecto de la etapa de labs, me pongo en el rol de MLOps Engineer, llevando a cabo el ciclo completo de un sistema de recomendación de películas, desde la preparación de datos hasta el desarrollo y despliegue de una API funcional con FastAPI. El objetivo final es crear un sistema que no solo ofrezca recomendaciones personalizadas, sino que también sea consumible por otros equipos dentro de una start-up de agregación de plataformas de streaming.

**Contexto y Descripción del Problema**

Como nuevo integrante en la start-up, me he enfrentado al desafío de crear un modelo de recomendación para películas con datos sin procesar. Los datos iniciales presentaban problemas como estructuras anidadas y valores faltantes, lo que dificultaba mi trabajo como Data Scientist. Sin embargo, mi tarea era comenzar desde cero, realizando tareas de ingeniería de datos y desarrollando un Minimum Viable Product (MVP) en un tiempo limitado.

**Rol a desarrollar**: Me he encargado de llevar los datos desde su estado bruto hasta una estructura utilizable, desarrollando un flujo de trabajo que incluye las siguientes fases:

* ETL (Extract, Transform, Load)
* Análisis Exploratorio de Datos (EDA)
* Modelamiento y despliegue de la API de recomendaciones

**Propuesta de Trabajo**

**Transformaciones Realizadas**

Las transformaciones que he llevado a cabo para preparar los datos de manera rápida y eficiente incluyen:

* **Desanidado de campos complejos**: Desanidé columnas como `belongs\_to\_collection` y `production\_companies` para poder realizar consultas eficaces en la API.
* **Manejo de valores nulos:**
  + Rellené valores nulos de `revenue` y `budget` con 0.
  + Eliminé filas con valores nulos en `release\_date`.
* **Formato de fechas**: Convertí las fechas al formato `AAAA-mm-dd` y extraje el año de lanzamiento en una nueva columna `release\_year`.
* **Cálculo de retorno de inversión**: Creé una columna `return` que calcula `revenue / budget` y asigna 0 si no se disponen de datos.
* **Eliminación de columnas innecesarias**: Eliminé `video`, `imdb\_id`, `adult`, `original\_title`, `poster\_path`, y `homepage` para reducir la complejidad del dataset.

**Desarrollo de la API**

He utilizado FastAPI para disponibilizar los datos transformados mediante endpoints clave. Las funciones desarrolladas incluyen:

1. **`cantidad\_filmaciones\_mes(mes)`:** Devuelve la cantidad de películas estrenadas en un mes dado.

2. `**cantidad\_filmaciones\_dia(dia)`:** Devuelve la cantidad de películas estrenadas en un día específico.

3. **`score\_titulo(titulo\_de\_la\_filmación)`:** Devuelve el título, año de estreno y score de popularidad de una película.

4. `**votos\_titulo(titulo\_de\_la\_filmación)`:** Devuelve la cantidad de votos y el promedio de puntuación de una película, solo si tiene al menos 2000 valoraciones.

5. `**get\_actor(nombre\_actor)`:** Devuelve el éxito del actor a través del retorno total, la cantidad de películas y el promedio de retorno por película.

6. **`get\_director(nombre\_director)`:** Devuelve el éxito del director, mostrando detalles como el retorno, la fecha de lanzamiento y las ganancias de cada película.

**Sistema de Recomendación**

El sistema de recomendación, desarrollado con técnicas de similitud, sugiere 5 películas similares a la ingresada por el usuario, ordenadas por score de similaridad. Esta funcionalidad está integrada en la API a través del endpoint:

* **`recomendacion(titulo)`:** Devuelve una lista con 5 películas recomendadas en orden descendente de similitud.

**Análisis Exploratorio de Datos (EDA)**

El análisis de los datos limpios incluyó la identificación de outliers, el análisis de correlaciones y la generación de visualizaciones como nubes de palabras para entender mejor la distribución de los títulos y patrones en los datos. Este análisis proporciona una base sólida para entrenar el modelo de recomendación.

**Despliegue**

El proyecto se ha desplegado utilizando Render. La API es accesible desde la web, y su funcionalidad se ha demostrado en un video de 7 minutos que resume las consultas y el modelo de recomendaciones.

**Video Demostrativo**

He creado un video que muestra el funcionamiento de la API, las consultas implementadas y una breve explicación del sistema de recomendación. Este video está disponible [enlace al video].

**Criterios de Evaluación**

1. **Código**: Prolijo, organizado y comentado donde es necesario. Uso adecuado de funciones y clases.

2. **Repositorio**: Archivos organizados y README detallado que permita entender el proyecto fácilmente.

3. **MVP Completo**: Cumple con las transformaciones y requerimientos descritos, con un sistema de recomendación funcional.

**Datos Utilizados**

- movies\_dataset.parquet

- credits.parquet

El proyecto incluye un diccionario de datos para facilitar la comprensión de las columnas.

¡Gracias por revisar el proyecto! Espero que encuentres útil y eficiente el sistema de recomendaciones de \*\*Proyecto\_CineSugerencias\*\*.