**DOCUMENTACION MODELO DE RECOMENDACIÓN DE LIBROS DATASET**

**PRUEBA TECNICA**

**CREADO POR: Adalberto Gonzalez Cuesta**

**PARA: Bancolombia**

**CIUDAD: Medellín, Co.**

**FECHA: 20/05/2024**

**AÑO:2024**

**OBJETIVO:** Se plantea la necesidad de crear un modelo de machine learning el cual debe ser entrenado con el Dataset de Amazon: <https://www.kaggle.com/datasets/mohamedbakhet/amazon-books-reviews>

El cual contiene la información de los libros que esta registrados en la biblioteca de libros de Google, con esta información se desea construir un modelo que tome los campos seleccionados por el usuario para saber de libros similares al de la entrada principal, (Ejemplo):

* Entrada: “El Señor de los Anillos: El Retorno del Rey”
* Salida: “Canción de Hielo y Fuego: Juego de Tronos”, “Las Crónicas de Narnia: El Príncipe Caspian”, “El Silmarillion”.

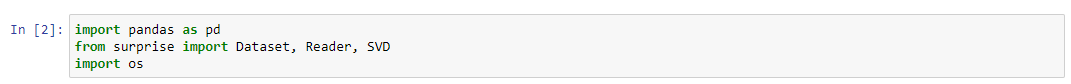
Al final el modelo debe de sugerir libros basados en las calificaciones de los usuarios para proporcionar recomendaciones personalizadas.

Nota: el modelo se creo tratando de utilizar valores singulares.

**LIBRERIAS UTILIZADAS:** os, pandas y surprise.

**PASO A PASO:**

Importación de las librerías necesarias.



Configuración de API, esto nos sirvió para bajar toda la data que vamos a utilizar para entrenar nuestro modelo (No es parte fundamental para el funcionamiento de nuestro modelo).



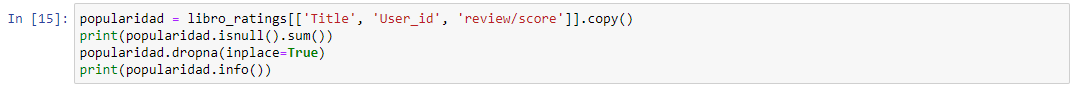
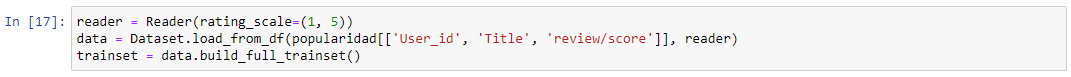
Carga de los datos de los archivos CSV utilizando la librería pandas.



Visualizamos la información contenida en los archivos CSV.



Preparamos todo lo que viene siendo el Dataset, en la cual especificamos las columnas que vamos a utilizar y le hacemos una copia de estos mismos, después validamos si hay valores nulos en cada unos de los campos, en caso de haber valores nulos los vamos a eliminar, y por último mostramos la data limpia.

Entrenamiento del modelo aquí estamos creando una instancia del algoritmo de SVD de la librería surprise.



Por ultimo se creó una función que recibe como parámetros el titulo de un libro, el entrenamiento del modelo, df que contiene las calificaciones de los libros y n que siempre va a ser 3, tratando de recomendar 3 libros similares al de la entrada la variable titulo.

