A movies/shows dataset with description is curated by pre-processing the Kaggle IMDb

Movies/Shows with Descriptions dataset [(https://www.kaggle.com/datasets/ishikajohari/imdb-datawith-descriptions)](https://www.kaggle.com/datasets/ishikajohari/imdb-data-with-descriptions)

The original dataset is pre-processed and is provided in 2 files - train.csv and test.csv. These files are provided to you each containing following columns:

* title: Title of the movie/show.
* description: Description of the movie/show.

You are required to train a text retrieval system using the train.csv file. And test the system using the test.csv file.

Your Tasks:

You are required to download the data sets, i.e., train.csv and test.csv. You must build a text retrieval system to find similar movies/shows based on the descriptions. You should systematically approach the problem by addressing the below tasks:

* Load the data sets and pre-process them to fit your requirements. You must use at least two pre-processing techniques.

* Design a text retrieval system using TF-IDF (with inverted file) algorithm.
* Find the top 3 movies/shows matches in the train.csv based on the descriptions provided in the test.csv.
* Improve the performance of the system using relevant metrics and approaches.

**The following guidelines is to be maintained**

* The use of any Python third-party package(s) is restricted to the following tasks:

* Loading the datasets. E.g., Pandas.
* Any necessary text pre-processing steps. E.g., Natural Language Toolkit, etc.
* Performing necessary calculations during the building of the system. E.g., NumPy.
* Calculating the performance of the system. E.g., Scikit Learn, Matplotlib, Plotly, etc.

* You should **NOT** use any third-party package for calculating TF-IDF (with inverted file).

* You should **ONLY** use the provided files, i.e., train.csv and test.csv for training/testing your system.