# IMDb\_score\_prediction

Data Source: https://www.kaggle.com/datasets/luiscorter/netflix-original-films-imdb-scores

Reference:Kaggle.com

# Prerequisites

Before you begin, ensure you have met the following requirements:

Python 3 installed on your machine.

# Required Python libraries: Pandas, NumPy, Scikit-Learn, Matplotlib, Seaborn.

You can install the necessary libraries using the following command:

# pip install pandas numpy scikit-learn matplotlib seaborn

# How to Run the Code

1)Download the dataset:

Download the IMDb dataset from https://www.kaggle.com/datasets/luiscorter/netflix-original-films-imdb-scores and place it in the data directory.

2)Run the Jupyter Notebook:

# jupyter notebook imdb\_score\_prediction.ipynb

Open the Jupyter Notebook in your browser and run each cell sequentially to see the data analysis, preprocessing, model training, and evaluation steps.

3)Alternatively, you can run the Python script directly:

# python imdb\_score\_prediction.py

This will execute the code in imdb\_score\_prediction.py and display the predicted IMDb scores for a sample of movies.

# Project Structure

data/: This directory contains the dataset files.

imdb\_score\_prediction.ipynb: Jupyter Notebook containing the code for data analysis, preprocessing, model training, and evaluation.

imdb\_score\_prediction.py: Python script implementing IMDb score prediction.

README.md: This file, providing instructions and information about the project.

# Dataset

The dataset used in this project contains information about movies, including features like IMDb rating, title, runtime, and more. The dataset is in CSV format and is located in the data directory.

# Methodology

# Data Loading

Load the dataset into a Pandas DataFrame.

# Data Exploration

Explore the dataset to gain insights and understanding of the features.

# Data Preprocessing

Clean the data by handling missing values and converting categorical variables into numerical representations.

# Feature Selection

Select relevant features for the prediction model.

# Model Selection

Choose appropriate machine learning algorithms (e.g., regression) for predicting IMDb scores.

# Model Training

Train the selected model using the training data.

# Model Evaluation

Evaluate the model's performance using appropriate metrics.

# Prediction

Use the trained model to predict IMDb scores for new data.

# DATA SOURCE WITH DESCRIPTION:

# Dataset Source

The dataset used in this project is sourced from IMDb, the popular online database of information related to films, television programs, and video games.

The dataset contains a comprehensive collection of movies with diverse attributes, making it ideal for predicting IMDb scores. Unfortunately, due to copyright and privacy concerns, we cannot provide a direct download link to the dataset.

However, you can obtain similar movie datasets from platforms like Kaggle (https://www.kaggle.com/datasets)

# Data Source: https://www.kaggle.com/datasets/luiscorter/netflix-original-films-imdb-scores

# git clone (https://github.com/MOHAMED-IBRAHIM-G/IMDb-score-prediction-.git)

# Dataset Description

The dataset used for this IMDb score prediction project includes the following features:

# Title

The title of the movie.

# Genre

The genre(s) of the movie, such as action, comedy, drama, etc.

# Runtime

The duration of the movie in minutes.

# Release Date

The date when the movie was released.

# Language

The language(s) spoken in the movie.

# IMDb Score

The target variable, representing the IMDb score of the movie (ranging from 1 to 10).