PROJECT OVERVIEW

Project Title: IMDb Score Prediction

Problem Definition: The problem is to develop a machine learning model that predicts IMDb scores of movies available on Films based on features like genre, premiere date, runtime, and language. The objective is to create a model that accurately estimates the popularity of movies, helping users discover highly rated films that match their preferences. This project involves data preprocessing, feature engineering, model selection, training, and evaluation.

Design:Dataset is taken from

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

Data Collection: Data has to be collected from various sources to form dataset. The dataset should contain some information about movies,including features like genre,premiere date,runtime,language and Imdb scores.

Data Preprocessing or cleaning: Clean the dataset by handling missing values,duplicate records and outliers.Ensure Consistency in data type and format.Once the data is cleaned it can be used to prepare for training.

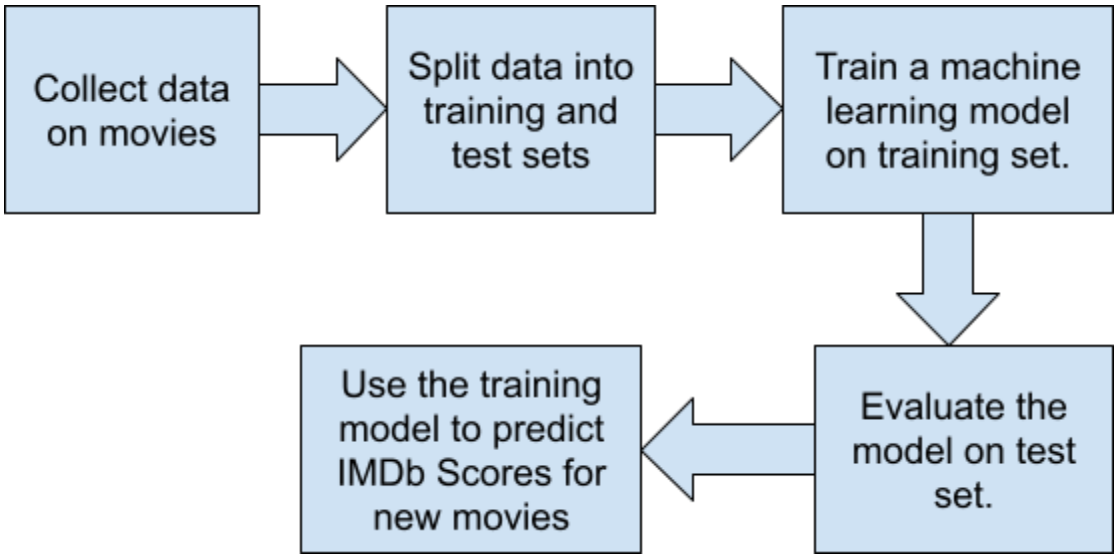
Feature Selection:
We need to select relevant features like genre,directors,actors,release year,runtime to predict IMDb Scores.

Model Selection:
We need to select an appropriate machine learning or regression model for predicting IMDb Scores.

Evaluation:
We will evaluate the model's performance using appropriate regression evaluation metrics that includes Mean Absolute Error (MAE),Mean Squared Error (MSE),Root Mean Squared Error (RMSE) and R-Squared (R2).

Iterative Improvement:
We will fine-tune the model parameters and explore techniques like feature engineering to enhance prediction accuracy.

FLOWCHART:


Collect data on movies
Split data into training and test sets
Train a machine learning model on training set.
Evaluate the model on test set.
Use the training model to predict IMDb Scores for new movies