

## **Problem definition of predicting IMDb scores :**

Predicting IMDb scores is a machine learning task that involves developing a model to estimate the rating that a movie or TV show is likely to receive on the IMDb platform. The problem can be defined as follows:

### **Problem Statement:**

Given a dataset of movies or TV shows with various features such as genre, director, actors, budget, and more, the task is to build a predictive model that can accurately forecast the IMDb score (rating) that a new, unseen movie or TV show is likely to receive based on its attributes.

### **Key Components:**

#### **1. Dataset:**

A collection of historical data on movies or TV shows, including their attributes (features) and their corresponding IMDb scores.

#### **2. Features:**

These are the input variables that the model uses for prediction. They can include factors like genre, director, actors, budget, release date, runtime, and more.

#### **3. Target Variable:**

The IMDb score is the target variable that the model aims to predict. IMDb scores typically range from 0 to 10, with higher scores indicating better user ratings.

#### **4. Machine Learning Model:**

You'll need to choose and develop an appropriate machine learning algorithm for regression since IMDb scores are continuous values. Common choices include linear regression, decision trees, random forests, or neural networks.

#### **5. Data Preprocessing:**

Cleaning and preparing the dataset is crucial. This involves handling missing data, encoding categorical features, and scaling or normalizing numerical features.

## **6. Model Training:**

Using a portion of the dataset to train the machine learning model. This step involves optimizing model parameters to achieve the best predictive performance.

## **7. Model Evaluation:**

Assessing the model's accuracy and performance using various metrics such as Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), or R-squared (R<sup>2</sup>) on a separate validation dataset.

## **8. Hyper parameter Tuning:**

Fine-tuning the model's hyper parameters to improve its predictive capabilities.

## **9. Model Deployment:**

Once satisfied with the model's performance, it can be deployed in a production environment to predict IMDb scores for new, incoming movies or TV shows.

## **10. Continuous Monitoring:**

Periodically updating the model and retraining it as new data becomes available to ensure its predictions remain accurate over time.

This problem definition forms the basis for developing a machine learning solution to predict IMDb scores, which can be valuable for filmmakers, studios, and streaming platforms to assess the potential success of their content.

**Done by:**

**Vulli Prem Kumar**