**Project Title: IMDb Score Prediction**

**Phase 1: Problem Definition and Design Thinking**

**Problem Definition**

The IMDb Score Prediction project aims to develop a machine learning model capable of predicting IMDb scores for movies available on Films. The prediction will be based on various movie attributes, including genre, premiere date, runtime, and language. The primary objective is to create a model that accurately estimates movie popularity, thereby assisting users in discovering highly rated films that align with their preferences. This project encompasses several essential steps, including data preprocessing, feature engineering, model selection, training, and evaluation.

**Design Thinking**

***1.*Data Source**

***Data Source:*** Utilize a dataset containing information about movies, including features such as genre, premiere date, runtime, language, and IMDb scores. The dataset will be obtained from the provided Kaggle link (Dataset Link: Netflix Original Films – IMDb Scores) .

**2.Data Preprocessing**

*Objective:* Clean and preprocess the dataset to ensure that it is suitable for model training.

**Method:**

***Handle missing values:*** Implement strategies such as imputation (e.g., mean or median imputation) for numerical features and mode imputation for categorical features.

***Convert categorical features:*** Utilize techniques like one-hot encoding or label encoding to transform categorical features into numerical representations.

***Normalize or standardize data:*** If necessary, scale numerical features to bring them to a common scale to aid model convergence.

**3.Feature Engineering**

*Objective:* Extract relevant features from the dataset that can contribute to predicting IMDb scores.

**Method:**

***Feature selection:*** Analyze the importance of each feature and select the most relevant ones for IMDb score prediction.

***Create new features:*** Generate additional features based on domain knowledge or insights gained during exploratory data analysis (e.g., seasonality from premiere date).

**4.Model Selection**

***Objective:*** Choose appropriate regression algorithms for predicting IMDb scores.

**Method:**

***Evaluate regression algorithms:*** Experiment with various regression models, such as Linear Regression, Random Forest Regressor, or Gradient Boosting Regressor.

Select the best-performing model based on evaluation metrics and model assumptions.

**5.Model Training**

***Objective:*** Train the selected regression model using the preprocessed data.

**Method:**

Split the dataset into training and testing sets to assess model performance.

Apply cross-validation techniques to ensure the model generalizes well.

**6.Evaluation**

***Objective:*** Evaluate the model's performance using regression metrics like Mean Absolute Error (MAE), Mean Squared Error (MSE), and R-squared (R2).

**Method:**

Calculate MAE and MSE to quantify the prediction error.

Use R2 to assess the proportion of variance in IMDb scores explained by the model.

By following this design thinking process, we will embark on a systematic journey to create a machine learning model that can accurately predict IMDb scores for movies. The success of this project will be determined by the model's ability to provide users with reliable recommendations and help them discover highly rated films that align with their preferences. Throughout the project, we will maintain flexibility to adapt to challenges and iterate on our approach as needed to achieve the best.

**Conclusion:**

The "IMDb Score Prediction" project aims to develop a robust machine learning model using a dataset sourced from Netflix original films. The objective is to accurately predict IMDb scores based on key movie attributes like genre, premiere date, runtime, and language. The project follows a systematic design thinking approach, encompassing data preprocessing, feature engineering, model selection, training, and evaluation. The success of the project will be gauged by the model's ability to assist users in discovering highly rated films that align with their preferences, enhancing the movie-watching experience.