**PREDICTING IMDB**

**SCORES**

**PHASE-1**

**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. IMDb scores are widely used to gauge the popularity and quality of movies, and being able to predict these scores accurately would be valuable for filmmakers, studios, and movie enthusiasts. The goal is to create a model that can predict the IMDb score of a movie based on various features such as cast, crew, genre, budget, release date, etc. This will help stakeholders in the movie industry make informed decisions and optimize their strategies.

**DESIGN THINKING**:

**Data Source**: Utilize a dataset containing information about movies, including features like genre, premiere date, runtime, language, and IMDb scores.

**Data Preprocessing:** Clean and preprocess the data, handle missing values, and convert categorical features into numerical representations.

**Feature Engineering**: Extract relevant features from the available data that could contribute to predicting IMDb scores.

**Model Selection**: Choose appropriate regression algorithms (e.g., Linear Regression, Random Forest Regressor) for predicting IMDb scores.

Model Training: Train the selected model using the pre-processed data.

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

**Empathize:**

* + Understand the needs and expectations of the stakeholders, such as filmmakers, studios, and movie enthusiasts.
  + Gather insights from IMDb users and industry experts to identify their pain points and desired features in a score predictor.
  + Explore existing models and tools for movie score prediction to understand their limitations and areas for improvement.

**Define:**

* + Clearly define the problem statement: "Develop an IMDb score predictor that accurately predicts the rating of a movie based on various features."
  + Identify the key features and data points that could influence the IMDb score, such as cast, crew, genre, budget, release date, etc.
  + Define the success criteria, such as achieving a certain level of accuracy in predicting IMDb scores.

**Ideate:**

* + Brainstorm potential approaches and algorithms for predicting IMDb scores, such as regression models, machine learning algorithms, or deep learning architectures.
  + Consider feature engineering techniques to extract meaningful information from the available data, such as text analysis for reviews or sentiment analysis.
  + Explore external data sources that could enhance the prediction accuracy, such as social media sentiment about movies or historical IMDb scores.

**Prototype:**

* + Develop a prototype IMDb score predictor using a chosen algorithm or approach.
  + Implement data preprocessing steps to clean and transform the input features into a suitable format for the model.
  + Split the dataset into training and testing sets to evaluate the model's performance.
  + Iterate on the prototype, fine-tuning the model and experimenting with different hyperparameters to improve accuracy.

**Test:**

* + Evaluate the model's performance using appropriate evaluation metrics, such as mean squared error (MSE) or R-squared.
  + Validate the model against a holdout dataset or perform cross-validation to assess its generalization capability.
  + Seek feedback from stakeholders and users, and incorporate their suggestions for further improvements.

**Implement:**

* + Develop a user-friendly interface or application that allows stakeholders to input movie features and obtain predicted IMDb scores.
  + Ensure the model's scalability and efficiency to handle a large number of predictions.
  + Deploy the IMDb score predictor in a production environment, making it accessible to the intended users.

**Learn and Iterate**:

* + Monitor and collect feedback from users to identify any issues or areas for enhancement.
  + Continuously update the model with new data and adapt to changing trends and user preferences.
  + Stay updated with new research and advancements in the field of movie rating prediction to incorporate innovative techniques.