**1. Data Collection:**

* Gathering a dataset of IMDb movies that includes features such as movie title, director, genre, release year, cast, budget, runtime, and any other relevant information. IMDb had provided the datasets for this purpose.

**2. Data Preprocessing:**

* Cleaning and preprocessing the data by handling missing values, removing duplicates, and converting categorical variables into a suitable format for modelling.

**3. Feature Engineering:**

* Extracting meaningful features from the data that might influence IMDb scores. features are the number of A-list actors in the cast, the director's previous success, and the movie's budget relative to its genre.

**4. Exploratory Data Analysis (EDA):**

* Performing EDA to understand the relationships between different features and IMDb scores. Visualizations and statistical analysis can helping us identify patterns and correlations.

**5. Model Selection:**

* Choosing an appropriate machine learning model for regression since IMDb scores are continuous values. Common choices of us is Linear Regression,

**6. Model Training:**

* Splitting the dataset into training and testing sets to train and evaluating the model. Ensuring the proper validation technique cross-validation to prevent overfitting.

**7. Feature Selection:**

* Using feature selection techniques for Linear Regression to identify the most influential factors in predicting IMDb scores.

**8. Evaluation Metrics:**

* Choosing appropriate evaluation metrics to assess the model's performance. Metrics for regression tasks is Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and R-squared (R²) score.

**9. Model Interpretability:**

* Making efforts to interpret the model's predictions. SHAP values or LIME helps explain why the model made a particular prediction.

**10. Continuous Improvement:**

* Regularly update the model with new IMDb data to keep it up-to-date and accurate. we refine our model by incorporating user feedback.