Phase 4

Development Part 2

In this part you will continue building your project.

Feature Engineering:

Feature engineering involves selecting and creating the right features that can help improve the predictive power of your model. In the context of IMDb score prediction, here are some feature ideas.

Creating a complete IMDb score prediction involves several steps and code that might be too extensive to provide in a single response. However, I can give you an outline and some sample code in Python to get you started. You can expand upon this foundation as needed:

Import Necessary libraries:

```
import pandas as pd
import numpy as np
from sklearn.model_selection impor
from sklearn.linear_model import L
from sklearn.metrics import mean_a
```

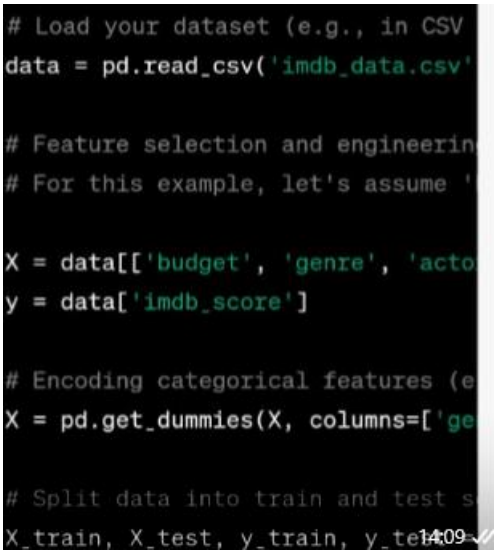
Loading and preparing data:

a.Genre Encoding: Convert movie genres into binary features (e.g., action, drama, comedy) to capture genre-specific trends.

b. Director and Actor Influence: Incorporate information about the directors and lead actors, such as the number of previous successful movies they've been involved in.

c. Release Date Features: Extract features like the year, month, or season of release, which could affect a movie's performance.

d. Budget and Box Office Data: Include data on a movie's production budget and box office earnings as potential features.

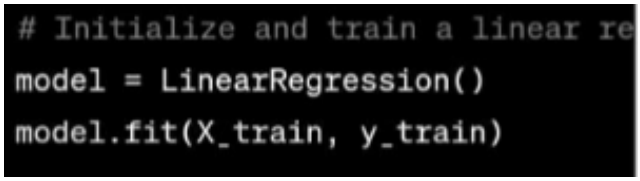


E.User and Critic Reviews: Utilize sentiment analysis on user reviews and critic reviews to generate sentiment-based features.

f. Runtime: The length of the movie could also be a factor in predicting IMDb scores.

Model Training:

Once you have your feature set, you can move on to model training. Common models for regression tasks like IMDb score prediction include linear regression, decision trees, random forests, and gradient boosting algorithms like XGBoost or LightGBM. Neural networks, particularly deep learning models, can also be used, such as a feedforward neural network or a recurrent neural network (RNN) if working with textual data.



Model Selection: Choosing a suitable algorithm based on the nature of your features and the problem complexity.

Training: Fit the chosen model on the training data. You may need to fine-tune hyperparameters.

Evaluation:

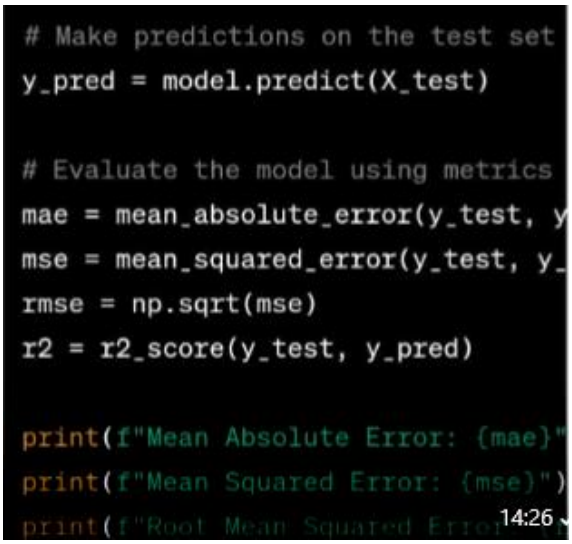
Evaluating your model's performance is critical to determine how well it predicts IMDb scores. Common evaluation metrics for regression tasks include:

a.Mean Absolute Error (MAE): Measures the average absolute difference between predicted and actual IMDb scores.

b. Mean Squared Error (MSE): Measures the average squared difference between predicted and actual scores.

c. Root Mean Squared Error (RMSE): The square root of MSE, providing error in the same units as IMDb scores.

d. R-squared (R2): Indicates the proportion of the variance in IMDb scores explained by the model.



Cross-Validation: Use techniques like k-fold cross-validation to ensure your model's performance isn't dependent on a specific data split.

Visualizations: Plot actual vs. predicted scores to gain insights into how well the model is performing.

By making adjustments to your model based on the evaluation results, and iterate on feature engineering and model training if necessary to improve performance.

Output:

Team Leader: SK.Nawab Sohail

Team Member: MOHAMMED Madhar

Team Member: MOHAMMED Abdul Raheem

Team Member: SK.Imran