**PREDICTING IMDB SCORES**

**Phase 4 Document submission**

Name : E. Tharun

Department : CSE-III

Register Number : 410121104058

NM ID : au410121104058

Domain : Applied Data Science

Project Title : Prediction IMDB Score

College : Adhi college of engineering and technology, Kanchipuram.

# Overview

This document consists of the implementation of feature engineering, model building and model evaluation in python.

Implementation of the above listed steps will be discussed brieﬂy with the necessary python code which was implemented and

executed successfully in google colab.

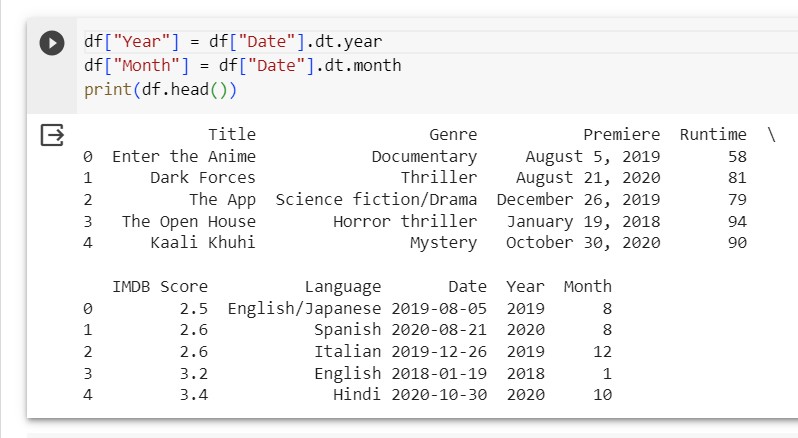
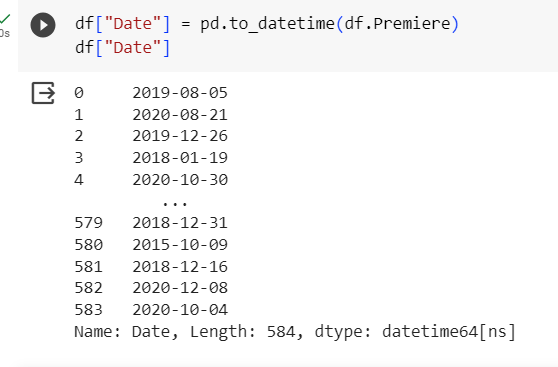
## Feature engineering

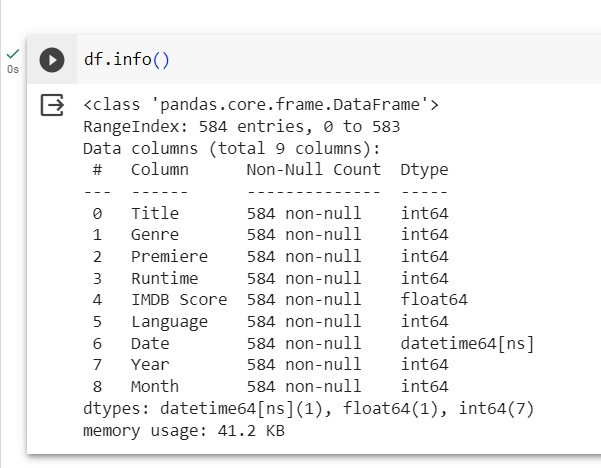
Feature engineering is the process of transforming raw data into features that are suitable for machine learning models. In other words, it is the process of selecting, extracting, and

transforming the most relevant features from the available data to build more accurate and eﬃcient machine learning models.

We have implemented certain Feature extraction process in the given dataset like extracting the columns ‘Date’ , ‘Year’ and ‘Month’ from the Premiere column that was already existing in the given dataset



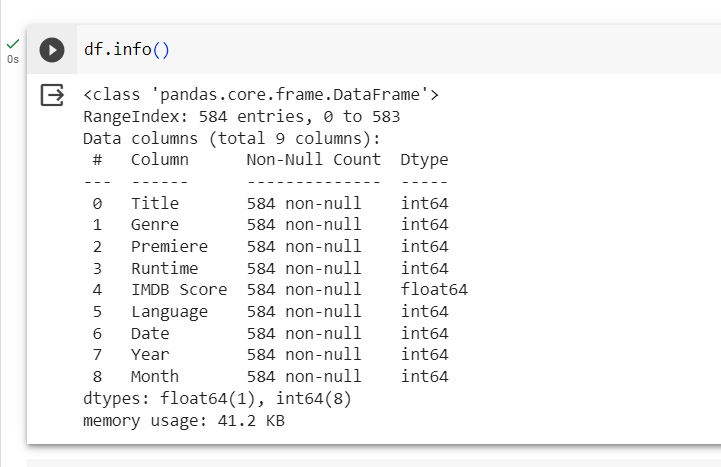




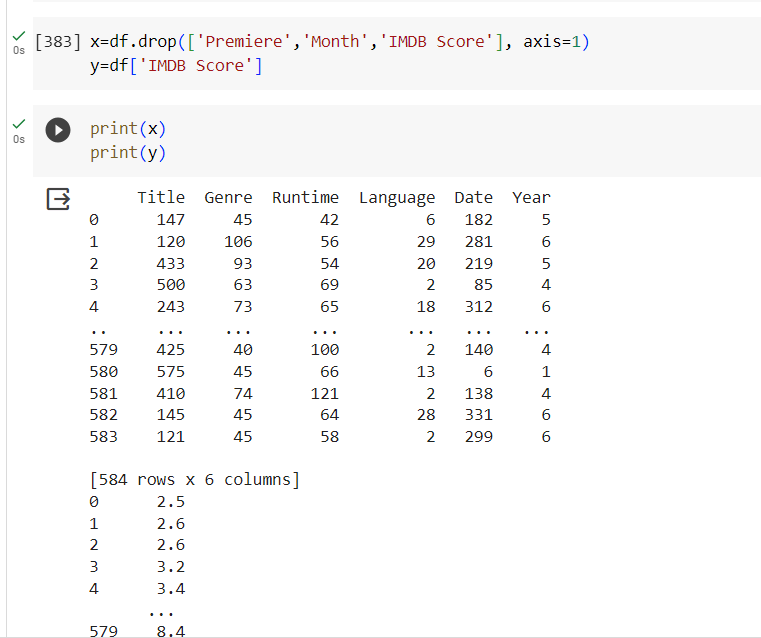
The Feature transformation was implemented by the use of Label encoder to convert the categorical values into numeric

values





The Feature selection is used to choose the most relevant features to include in the model while eliminating irrelevant or redundant ones. Therefore, we have excluded ‘Premiere’ and ‘Month’.



Splitting of data

- The dataset is divided into training and test sets. The training set is used to train the model and the test set is used to evaluate the model's generalization performance.

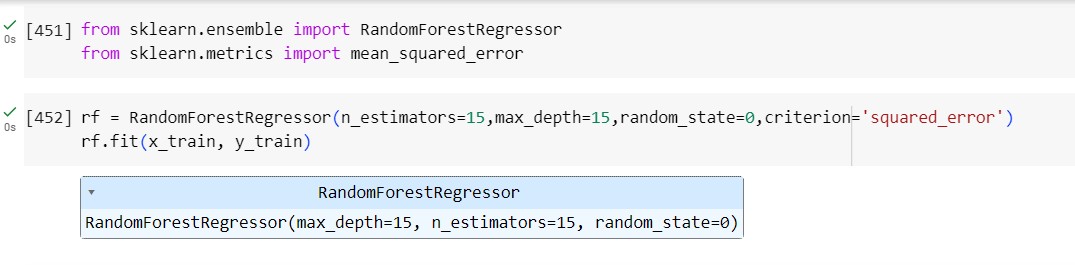


## Model Building

By choosing a machine learning algorithm a model is build for Predicting the IMDb Scores for the given dataset

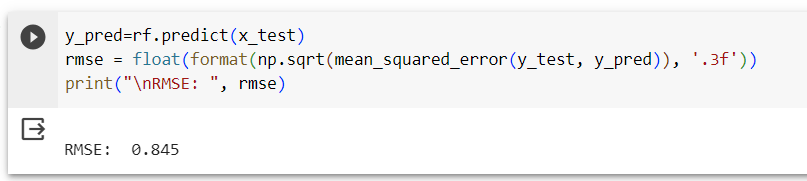
Model Training

Train the model on the training dataset. The model learns the patterns and relationships in the data during this phase.



Model Validation

Validate the model on the test set to assess its generalization performance. This step ensures that the model can make accurate predictions on unseen data.



Model Evaluation

Evaluation metrics for regression models are used to assess the performance of models that predict continuous numeric values.

These metrics help to understand how well the regression model is making predictions and are crucial for model selection, hyperparameter tuning, and comparing diﬀerent regression algorithms. Here are some common evaluation metrics for

regression models:

1. Mean Absolute Error (MAE):

* Measures the average absolute diﬀerence between actual and predicted values.
* Calculation: (1/n) Σ |actual - predicted|

1. Mean Squared Error (MSE):

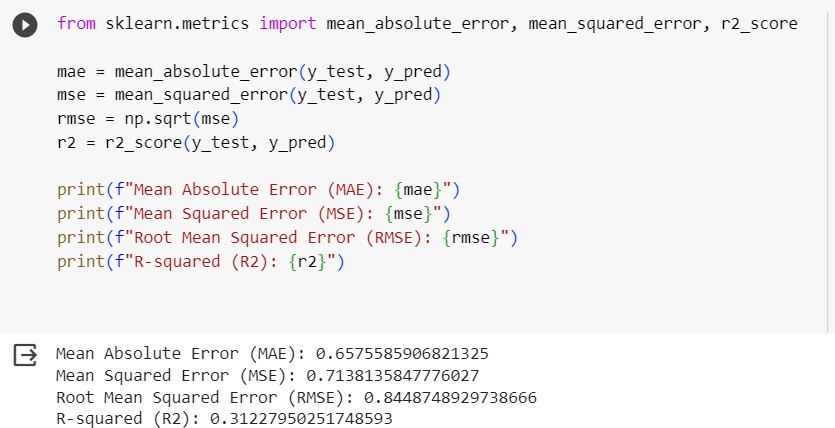
* Measures the average of the squared diﬀerences between actual and predicted values.
* Calculation: (1/n) Σ (actual - predicted)^2

1. Root Mean Squared Error (RMSE)\*\*:

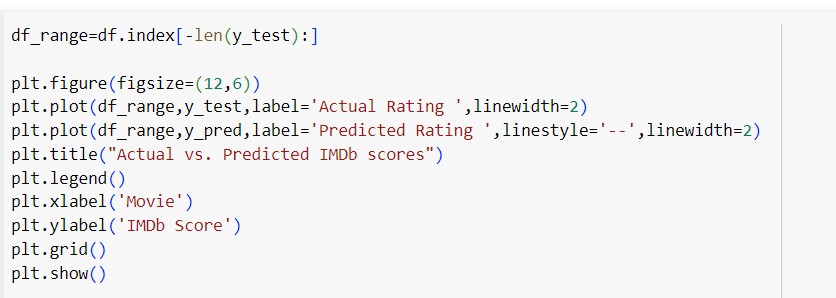
* It is the square root of the mean squared error.
* Calculation: √MSE

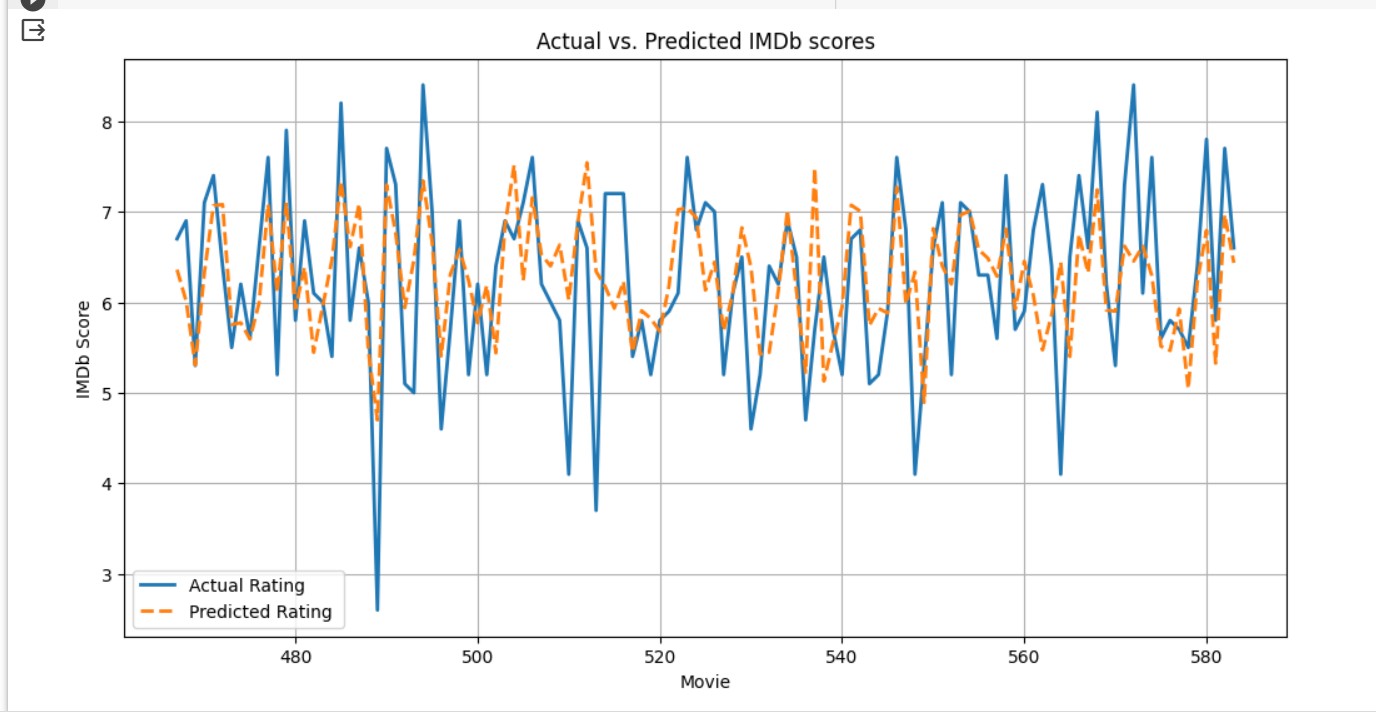
1. \*\*R-squared (R2) Score\*\*:

* Measures the proportion of the variance in the dependent variable that is predictable from the independent variables.
* Calculation: 1 - (MSE(model) / MSE(mean))



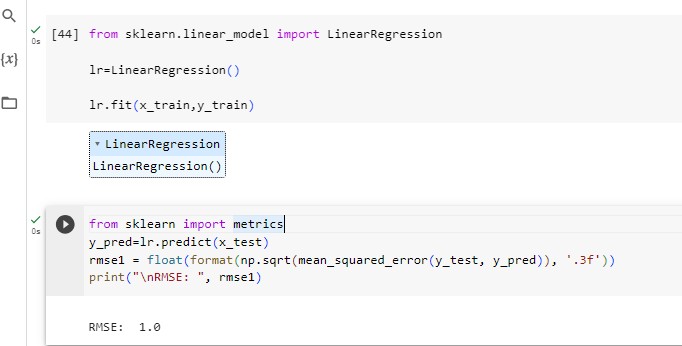
Visualization of Random Forest Regressor

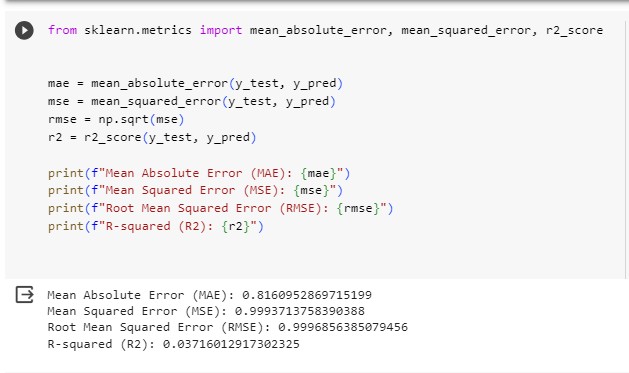




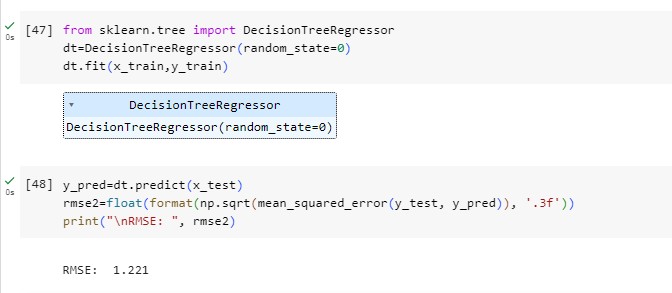
# Other Models we built

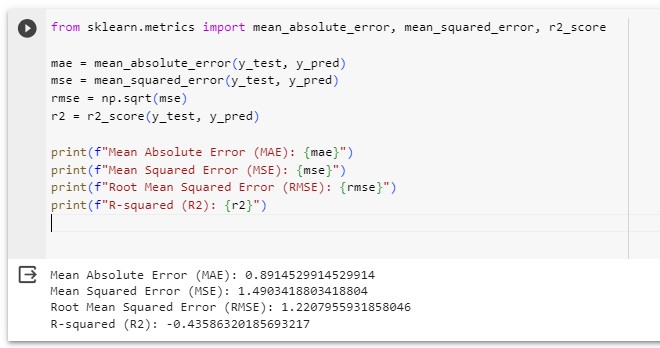
* 1. Linear Regression



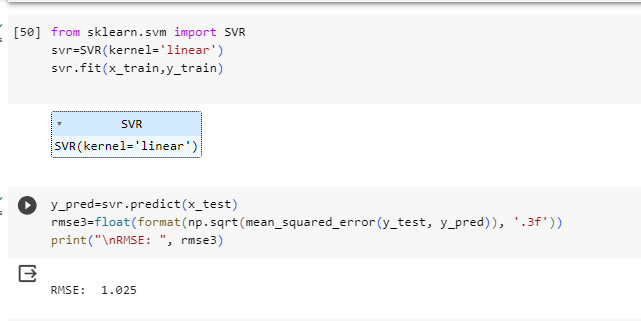


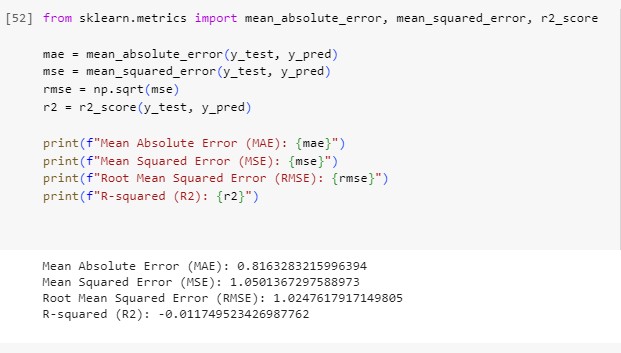
* 1. Decision Tree



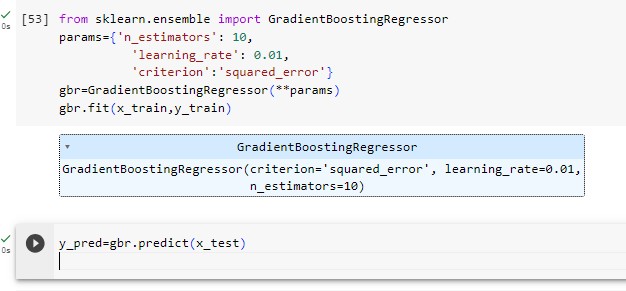


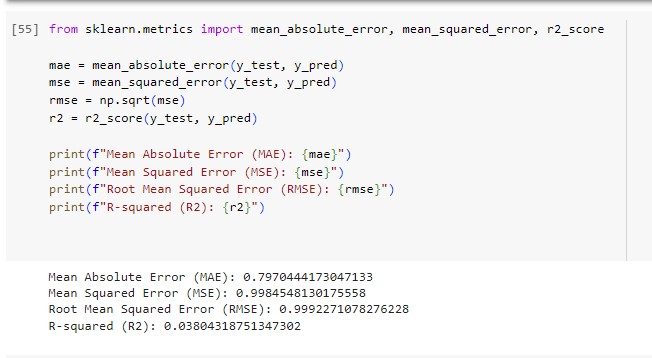
* 1. Support Vector Regressor



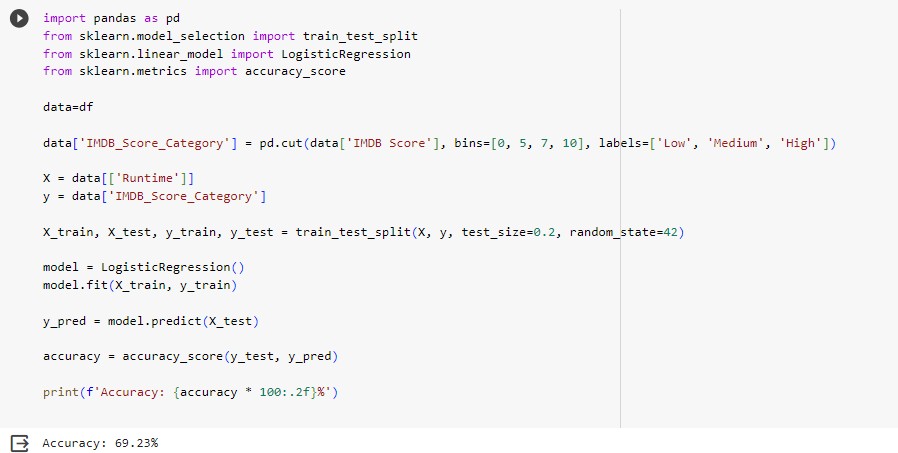


* 1. Gradient Boosting Regressor





* 1. Logistic Regression



* 1. Polynomial Regression

