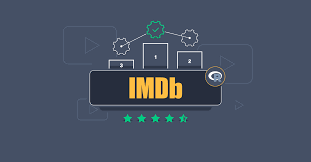
**Predicting IMDb Scores Using Machine Learning**

TEAM MEMBER

732521104054 : THIRISHA P

**Phase 2 Submission Document**

**Project :** Predicting IMDb Scores



Introduction:

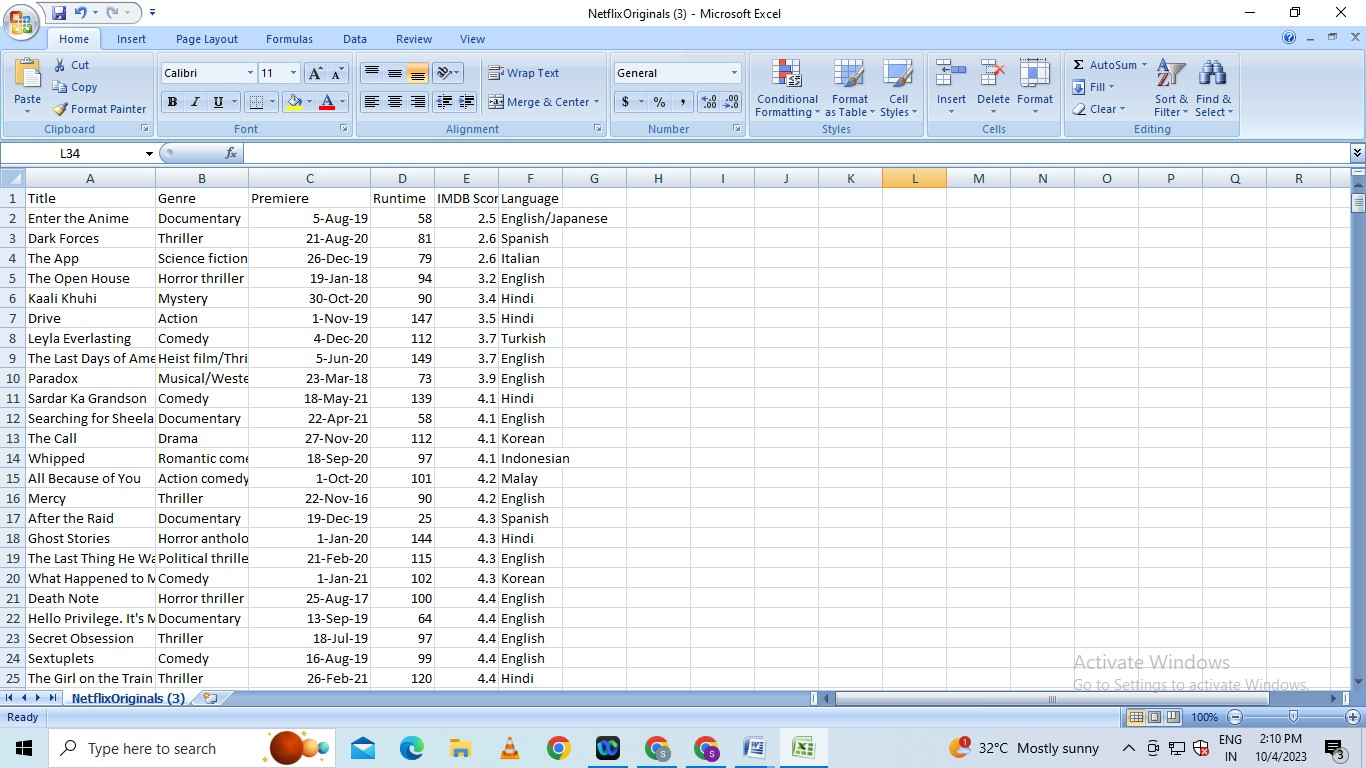
* Predicting IMDb scores for movies or TV shows typically involves using machine learning models and features such as cast, crew, genre, user reviews, and more. You can use regression algorithms to build a predictive model.
* The quality of your predictions depends on the quality and quantity of data, as well as the choice of features and model.
* In this project , we will explore advanced regression techniques to enhance the accuracy and robustness of IMDb scores prediction models.
* Highlight the limitations of traditional linear regression models in capturing complex relationships.
* Emphasize the need for advanced regression techniques like Gradient Boosting and Neural Networks to enchance prediction accuracy.

**Content For Project Phase 2 :**

Consider exploring advanced regression technique like Gradient Boosting or Neural Networks for improved Prediction accuracy.

**Data Source :**

A Good Data for Predicting IMDb Scores using machine learning model should be Accurate , complete , accessible

**Dataset Link : (**[**https://www.kaggle.com/datasets/luiscorter/netflix-original-films-imdb-scores**](https://www.kaggle.com/datasets/luiscorter/netflix-original-films-imdb-scores))

**Model Evaluation and Selection:**

* Split the dataset into training and testing sets.
* Evaluate models using appropriate metrics (e.g., Mean Absolute Error, Mean Squared Error, R-squared) to assess their performance.
* Use cross-validation techniques to tune hyperparameters and ensure model stability. Compare the results with traditional linear regression models to highlight improvements.
* Select the best-performing model for further analysis.

**Model Interpretability:**

* Explain how to interpret feature importance from Gradient Boosting and Neural Networks.
* Discuss the insights gained from feature importance analysis and their relevance to IDMb scores prediction.
* Interpret feature importance from ensemble models like Random Forest and Gradient
* Boosting to understand the factors influencing IDMb.

**Deployment and Prediction:**

* Deploy the chosen regression model to predict IDMb.
* Develop a user-friendly interface for users to input property features and receive IDMb scores predictions.

**Program:**

**IDMb Score Prediction**

import pandas as pd

import numpy as np

import seaborn as sns

import matplotlib.pyplot as plt

from sklearn.model\_selection import train\_test\_split

from sklearn.preprocessing import StandardScaler

from sklearn.metrics import r2\_score,mean\_absolute\_error, mean\_squared\_error

from sklearn.linear\_model import LinearRegression

from sklearn.linear\_model import Lasso

from sklearn.ensemble import RandomForestRegressor

from sklearn.svm import SVR

import warnings

warnings.filterwarnings("ignore")

warnings.warn(" A NumPy version>={np\_minversion) and <{np\_maxversion}")

dataset = pd.read\_csv("C:/ibm:/Netflix.csv”)

**Model 1 - Linear Regression**

**In [1]:**

model\_Ir-LinearRegression()

**In [2]:**

model\_Ir.fit(X\_train\_scal, Y\_train)

**Out[2]:**

Linear Regression

Linear Regression()

**Predicting Prices**

In [3]:

Prediction= model\_Ir.predict(X\_test\_scal)

**Evaluation of Predicted Data**

In [4]:

plt.figure(figsize=(12,6))

plt.plot(np.arange(len(Y\_test)), Y\_test, label=Actual Trend)

plt.plot(np.arange(len(Y\_test)), Prediction1, label="Predicted Trend)

plt.xlabel(”Data”)

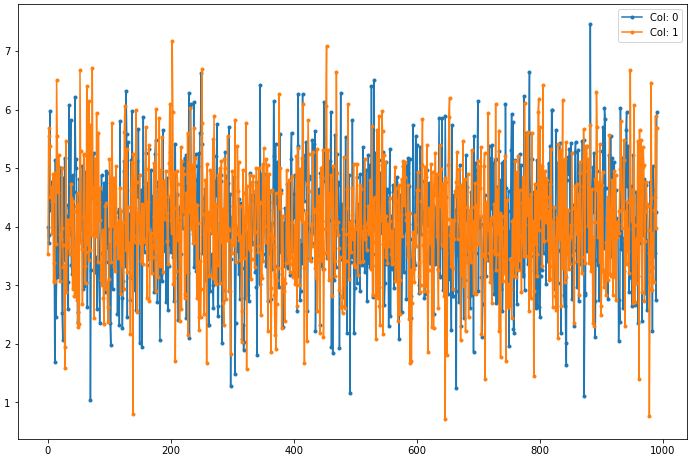
plt.ylabel("Trend”)

plt.legend()

plt.title(“Actual vs Predicted”)

**Out[4]:**

Text(0.5, 1.0, 'Actual vs Predicted')

Actual vs Predicted

**Conclusion and Future Work (Phase 2):**

**Project Conclusion:**

In the Phase 2 conclusion, we will summarize the key findings and insights from the advanced regression techniques. We will reiterate the impact of these techniques on improving the accuracy and robustness of IDMb score predictions.

Future Work: We will discuss potential avenues for future work, such as incorporating additional data sources (e.g., real-time economic indicators), exploring deep learning models for prediction, or expanding the project into a web application with more features and interactivity.