
PROJECT
IMDb Score Prediction

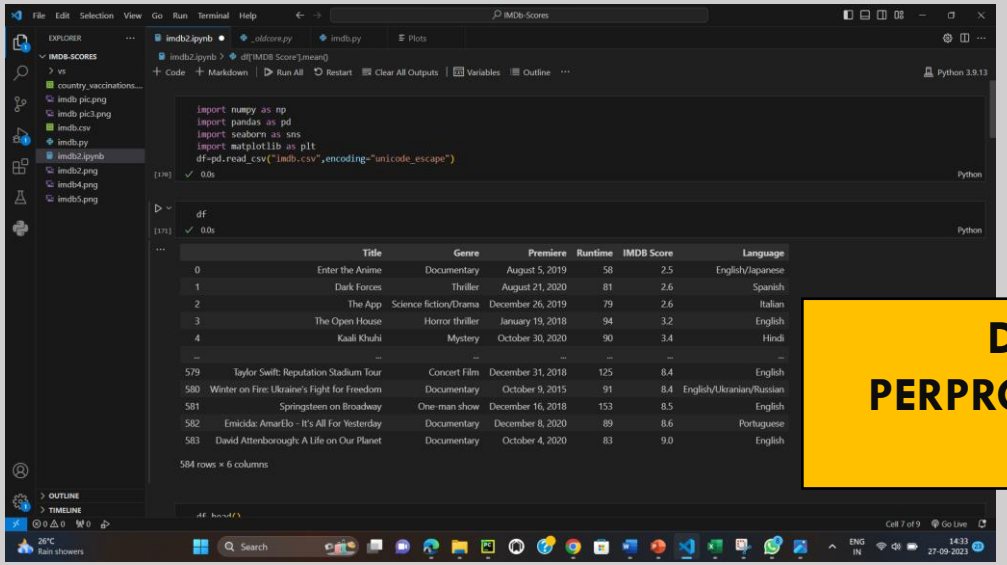
Develop a machine learning model to predict the IMDb Scores of the movies available on Films Based in their genre, premiere date, runtime and language . The model aims to accurately finds the popularity of the movies to the assist users in discovering highly rated films that align with their preferences

Design the project based on :

* Data Source
* Data Preprocessing
* Feature Engineering
* Model Selection
* Model Training
* Evaluation

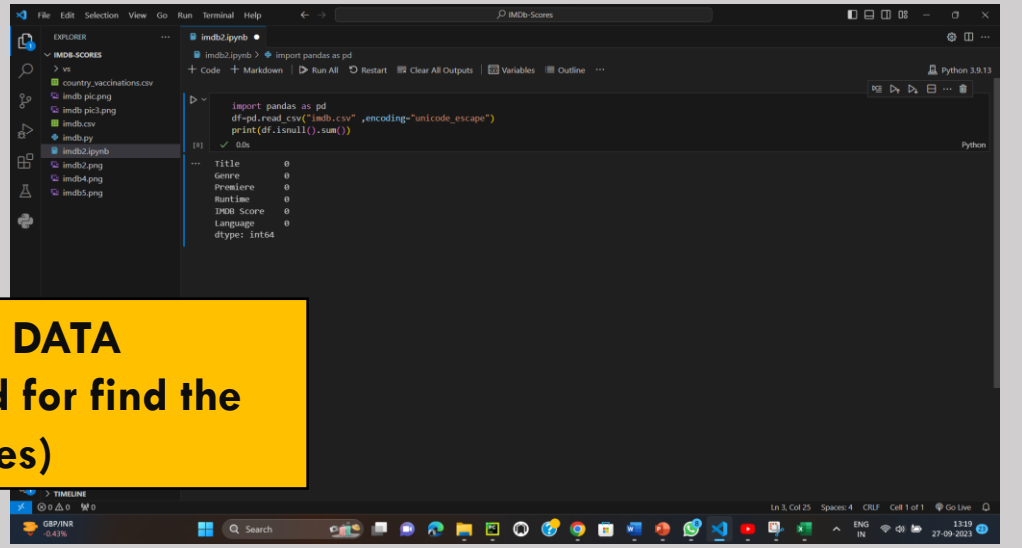
Main algorithms & ML are :

linear Regression ,Random Forest Regression to Predict IMDb Scores .Train the selected model using preprocessing data Regression metrics like MAE ,MSE and R-squared

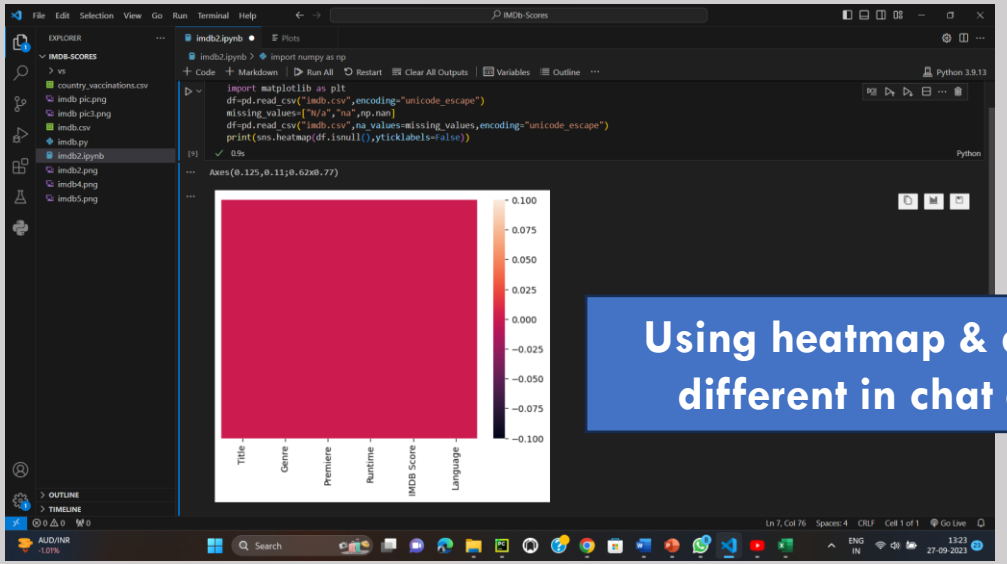


	Title	Genre	Premiere	Runtime	IMDB Score	Language
0	Enter the Anime	Documentary	August 5, 2019	58	2.5	English/Japanese
1	Dark Forces	Thriller	August 21, 2020	81	2.6	Spanish
2	The App	Science fiction/Drama	December 26, 2019	79	2.6	Italian
3	The Open House	Horror thriller	January 19, 2018	94	3.2	English
4	Kaali Khuhi	Mystery	October 30, 2020	90	3.4	Hindi
...	...	...	...	...	...	...
579	Taylor Swift: Reputation Stadium Tour	Concert Film	December 31, 2018	125	8.4	English
580	Winter on Fire: Ukraine's Fight for Freedom	Documentary	October 9, 2015	91	8.4	English/Ukranian/Russian
581	Springsteen on Broadway	One-man show	December 16, 2018	153	8.5	English
582	Emicida: AmarElo - It's All For Yesterday	Documentary	December 8, 2020	89	8.6	Portuguese
583	David Attenborough: A Life on Our Planet	Documentary	October 4, 2020	83	9.0	English

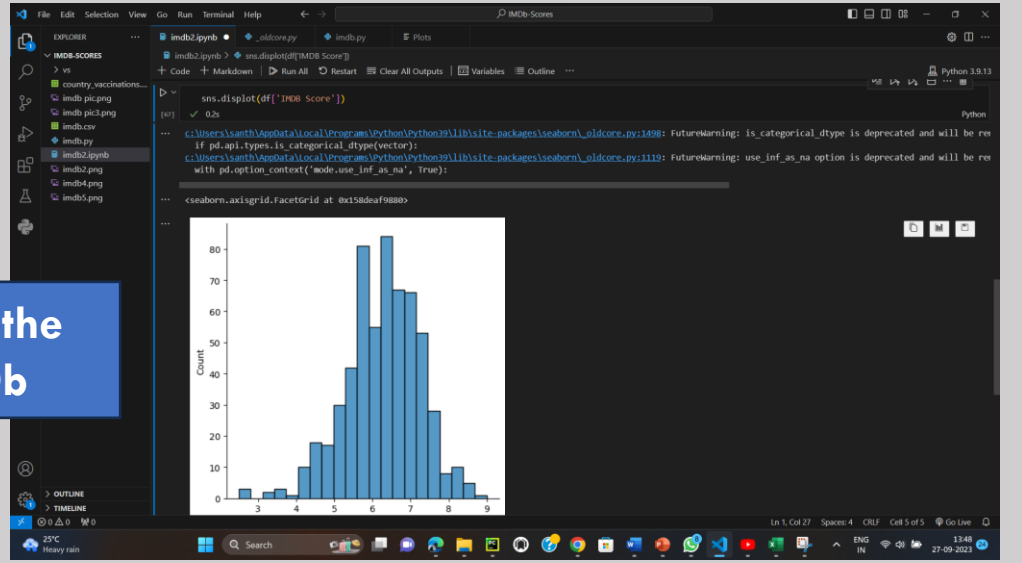
584 rows × 6 columns



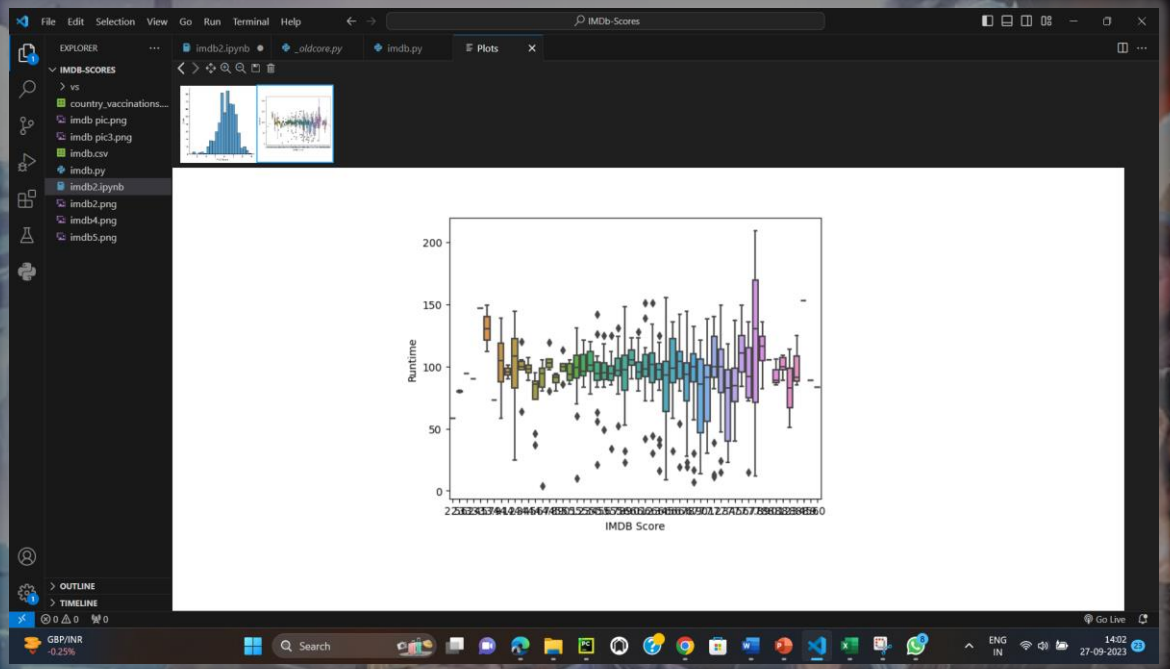
DATA SOURCE & DATA PERPROCESSING (used for find the missing values)

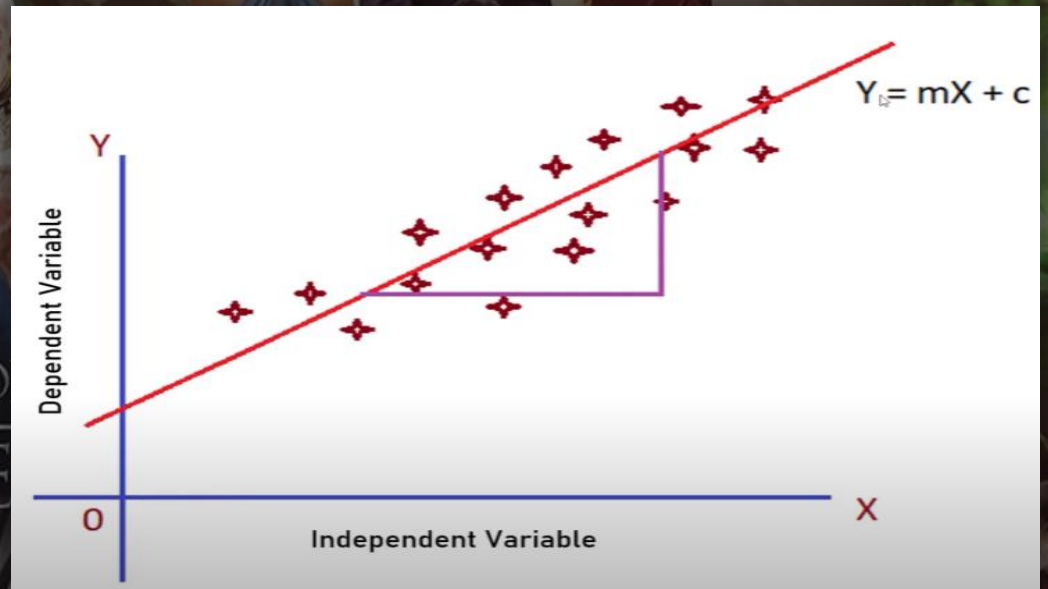


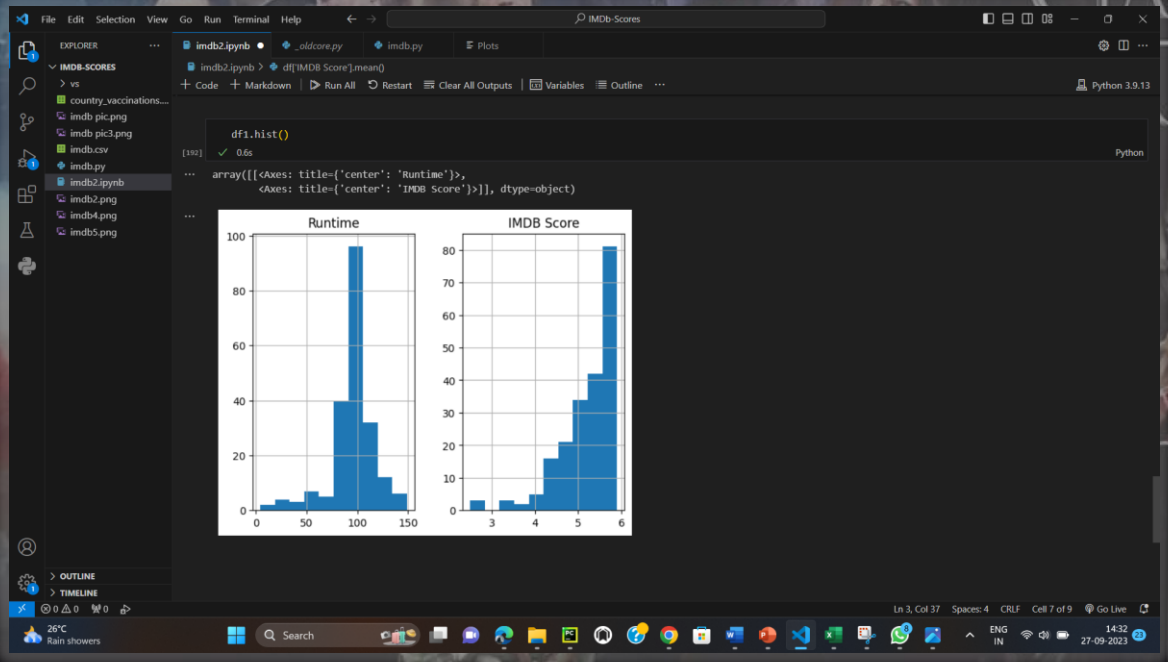
Using heatmap & displot can find the different in chat and predict IMDb



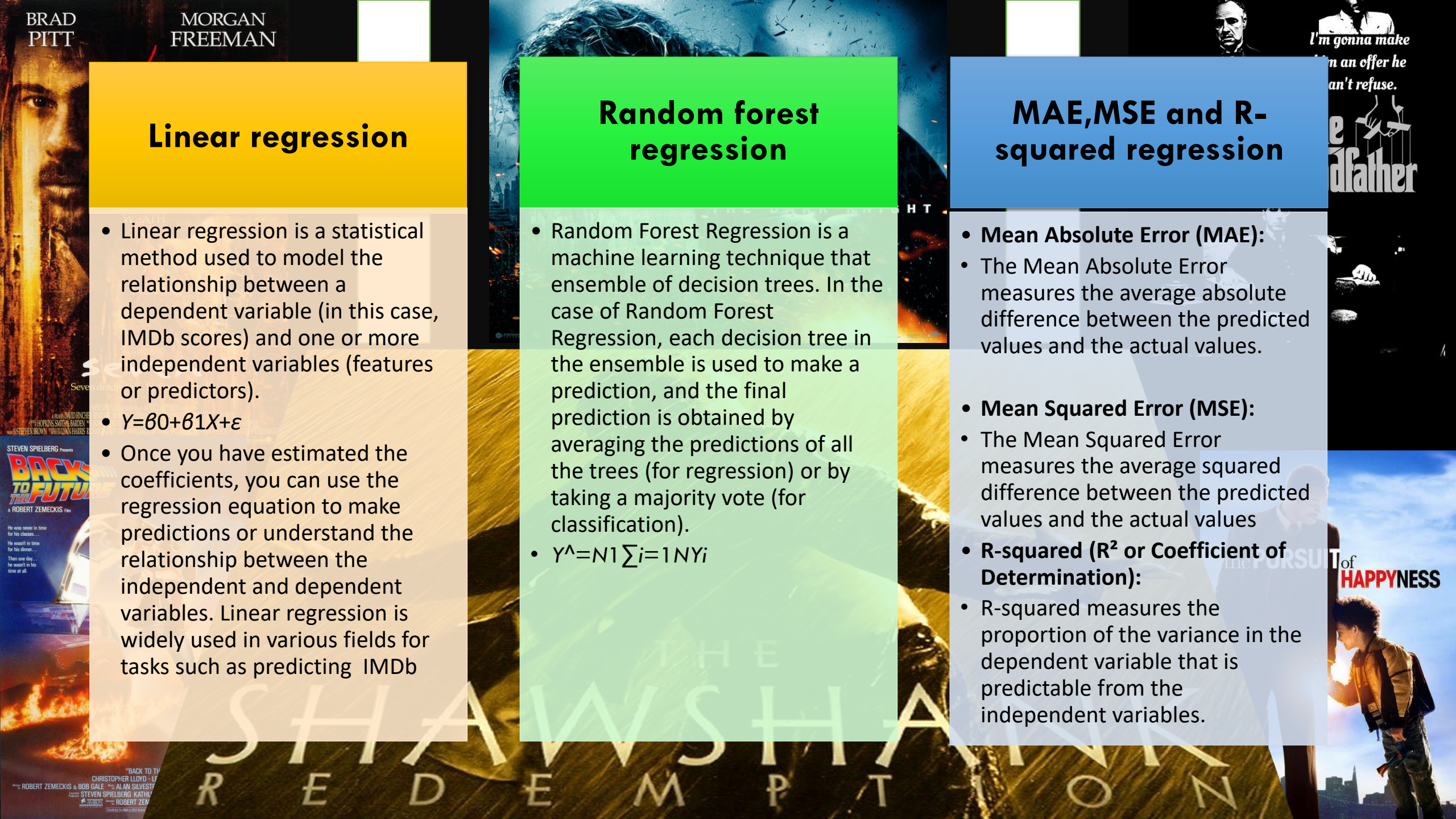
The model is selected and train by the machine learning by using linear regression & random forest regression in python







The above graph explain the training of the model in regression and using X and Y in the graph we can predict the value of the m & c it is preprocessing data in dataset and After we can evaluate the predicted Scores using MAE and MSE or R-squared



Linear regression

- Linear regression is a statistical method used to model the relationship between a dependent variable (in this case, IMDb scores) and one or more independent variables (features or predictors).
- $Y=\beta 0+\beta 1X+\varepsilon$
- Once you have estimated the coefficients, you can use the regression equation to make predictions or understand the relationship between the independent and dependent variables. Linear regression is widely used in various fields for tasks such as predicting IMDb

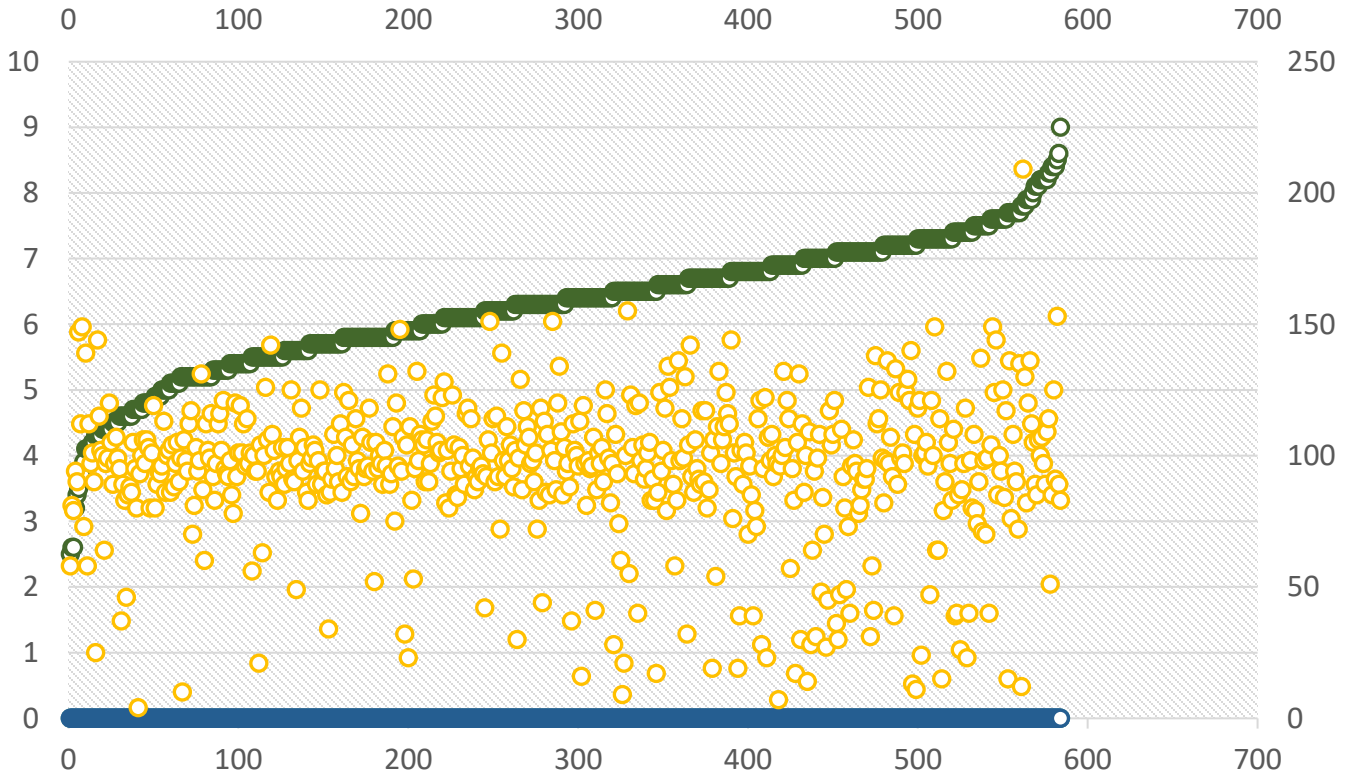
Random forest regression

- Random Forest Regression is a machine learning technique that ensemble of decision trees. In the case of Random Forest Regression, each decision tree in the ensemble is used to make a prediction, and the final prediction is obtained by averaging the predictions of all the trees (for regression) or by taking a majority vote (for classification).
- $Y\hat{}=N1\sum i=1NYi$

MAE,MSE and R-squared regression

- **Mean Absolute Error (MAE):**
- The Mean Absolute Error measures the average absolute difference between the predicted values and the actual values.

- **Mean Squared Error (MSE):**
- The Mean Squared Error measures the average squared difference between the predicted values and the actual values
- **R-squared (R^2 or Coefficient of Determination):**
- R-squared measures the proportion of the variance in the dependent variable that is predictable from the independent variables.



IMDb Scores

REGRESSION MODEL OF IMDb SCORE PREDICTION