2025

SARVESH PRATAP YADAV

**Indian movies IMDb rating predictions using Machine Learning Model.**

**Problem Statement –** The prediction of success of movie with good accuracy is needed

in the film industry which helps different people working in the film industry manly for

the investors.

The aim of this project is to create a movie rating prediction system which will predict a

rating of the movie which will use to determine that should he gives his time on a

particular movie or should go for any other movie. Instead of getting a review from a

particular person, we will provide it by using algorithms of machine learning and dataset

from previous experience of users.

**Steps to be taken in the Project is sub-divided into the following sections.**

**These are:**

 Loading necessary libraries such as numpy , pandas , sklearn etc.

 Loading the dataset as CSV file and showing first 10 rows.

 Calculate statistical values and round them up to 3 decimal places.

 Checking for null values and return their sum of numbers of true values in each

column.

 Drop all the null values from data.

 Extracting all information about data.

 Checking Shape of Data.

 Replacing string values available in column values by changing their datatypes.

 Checking for the null values in dataset.

 Drop the unnecessary columns from dataset.

 Visualization of Sales by different source of Indian movies IMDb data using

Python data visualization.

 Data preprocessing or (Data cleaning) performed by the one hot encoding in

this process we change categorical data into numerical data and the

technique is called feature Engineering.

 Splitting the cleaned data into dependent and independent variables.

 Splitting the data into train and test sets with train\_test\_split using sklearn

library.

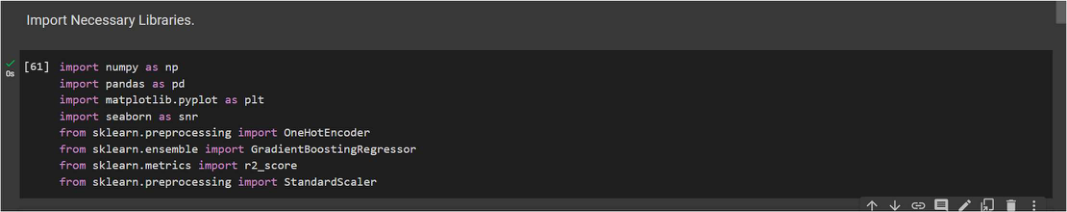
 Import different kind of Regression Models and Train that model with the

help of .fit().

 Predicting the trained models and then checking their accuracy of the model

using accuracy score.

 Then recall the train\_test\_split and split the data into training and testing set



2025

SARVESH PRATAP YADAV

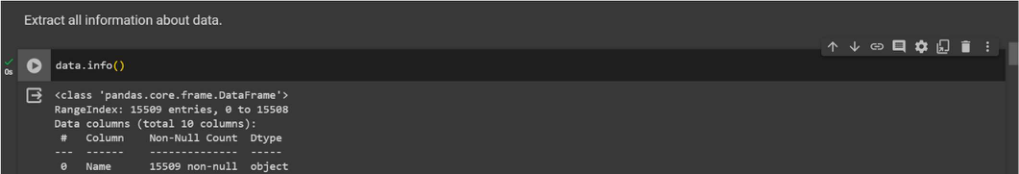
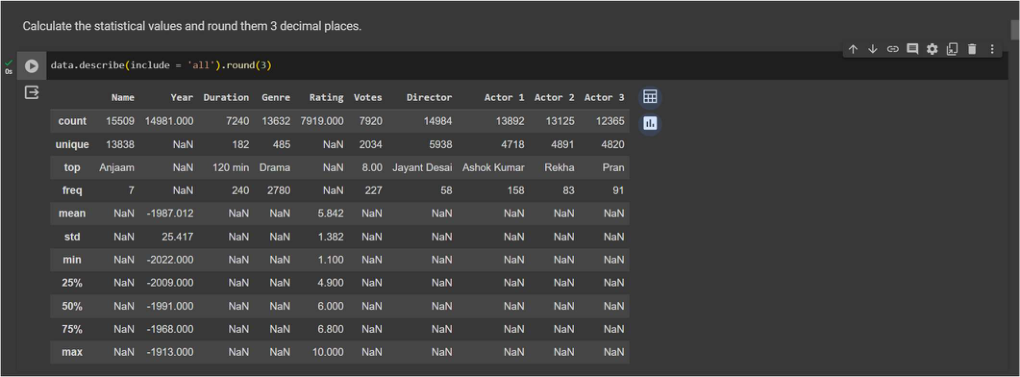
with different models.

 Then predicting the trained models and checking the accuracy of model and

print the accuracy.

**Step-1 –** Loading Necessary Libraries used in machine learning.

**Step-2 –** Loading the dataset as csv file and showing first ten rows.



SARVESH PRATAP YADAV

**Step-3 –** Calculate statistical values and round them up to 3 decimal places.

**Step-4 –** Extracting all information about data.

**Step-5 –** Checking Shape of data.



2025

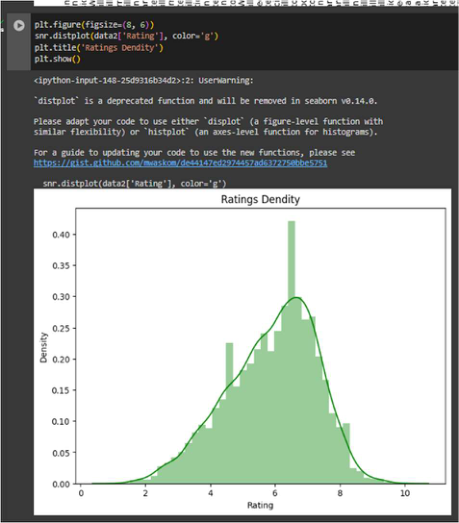
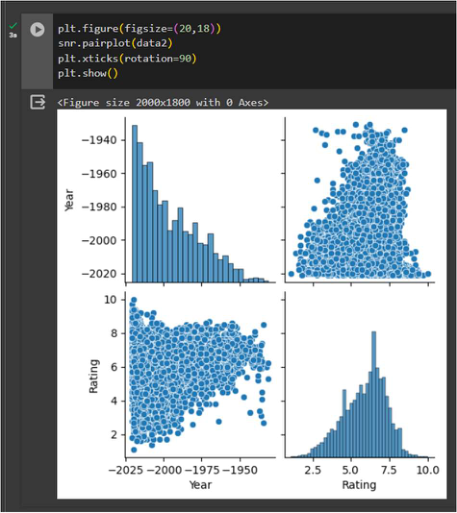
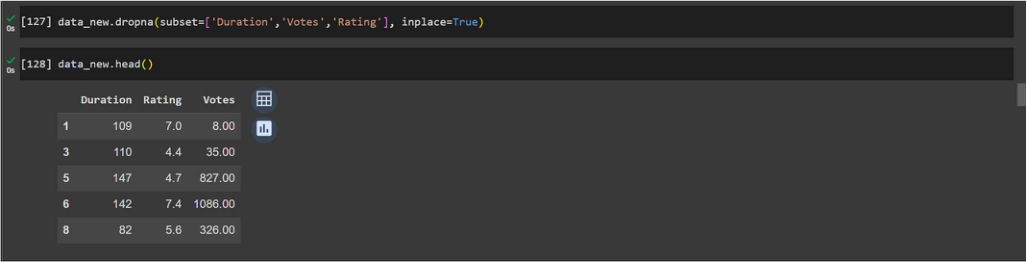
SARVESH PRATAP YADAV

**Step-6 –** Drop the unnecessary columns from dataset.

**Step-7 –** Replacing string values available in column values by changing their   
 datatypes.

**Step-6 –** Checking for null values and return their sum of numbers of true values in each

column.

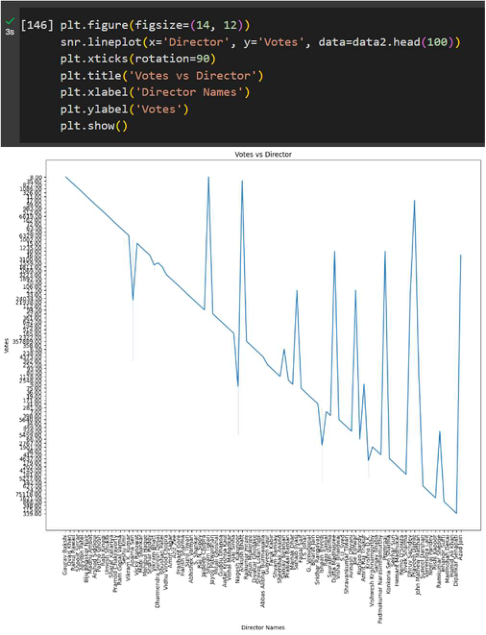
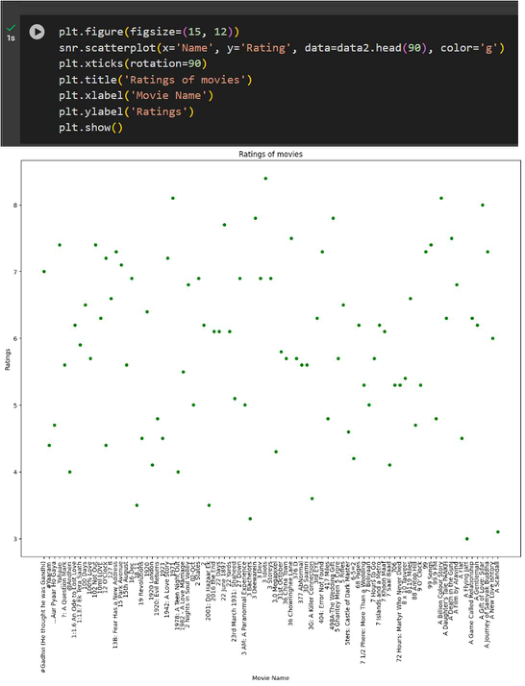


SARVESH PRATAP YADAV

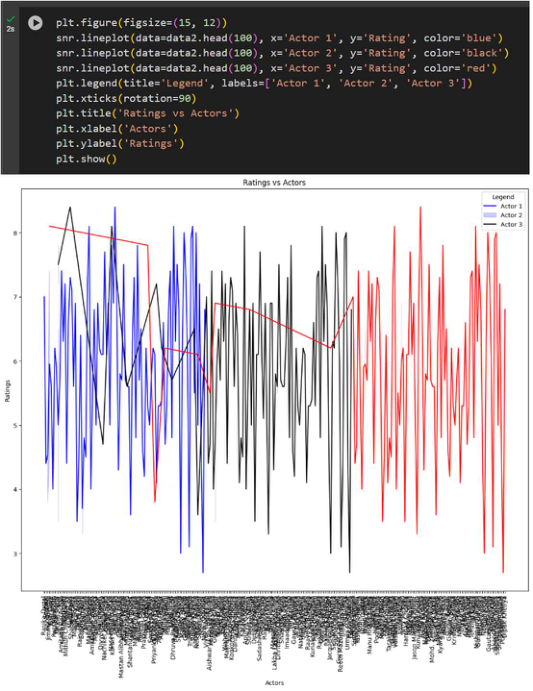
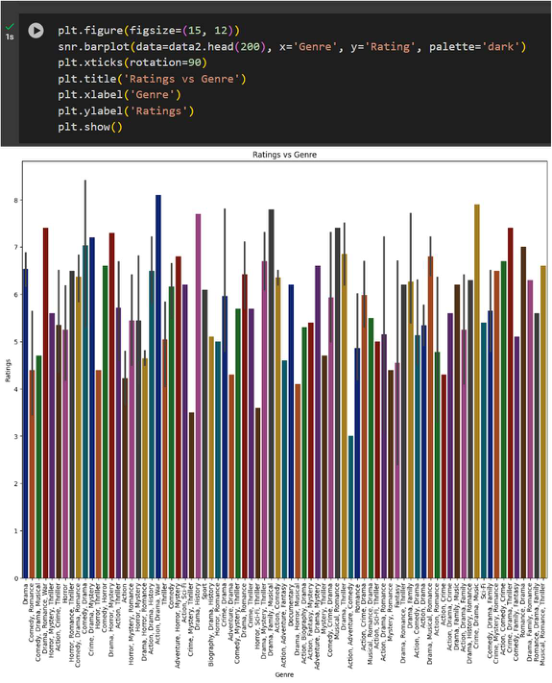
**Step-7 –** Drop all the null values from data.

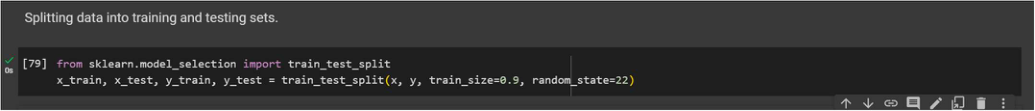
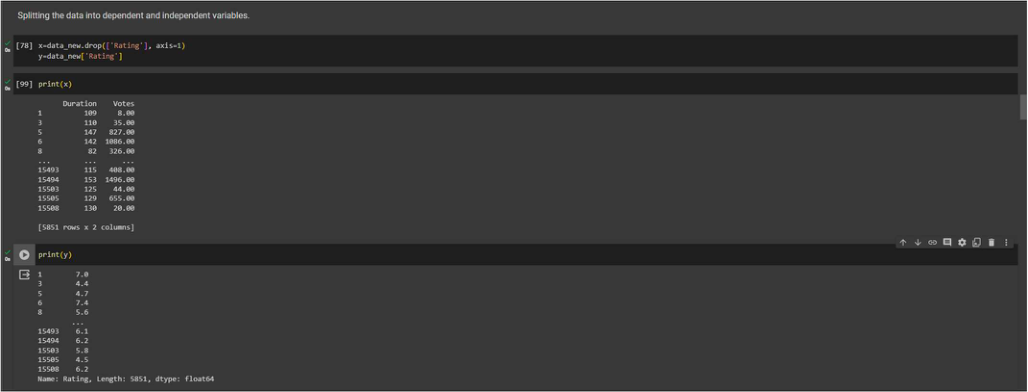
**Step-8 –** Visualization of Sales by different source of Indian movies IMDb data

using Python data visualization.



SARVESH PRATAP YADAV





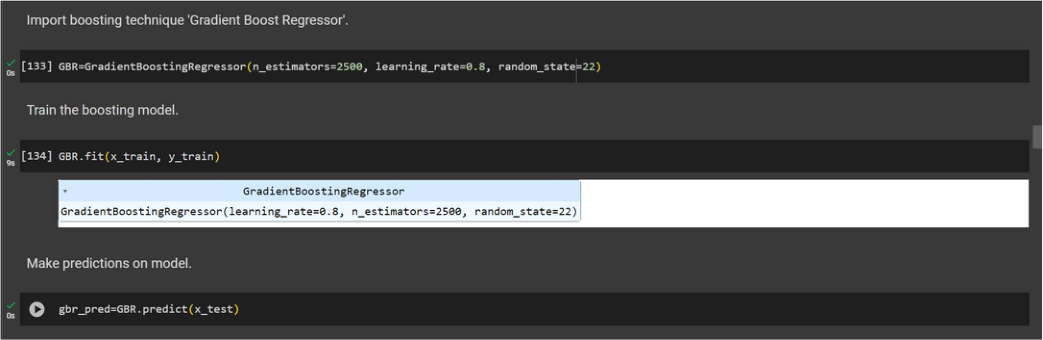
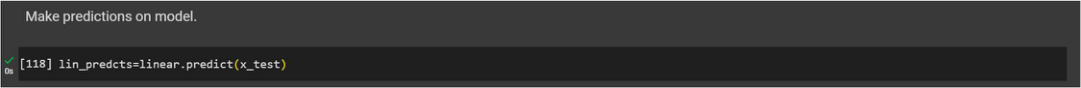
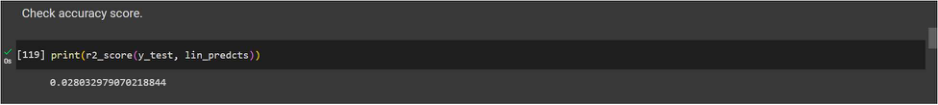
SARVESH PRATAP YADAV

**Step-9 –** Splitting the data into dependent and independent variables.

**Step-10 –** Splitting the data into training and testing sets.

**Step-11 –** Import first machine learning model ‘Linear Regression’.

**Step-12 –** Train the model.



SARVESH PRATAP YADAV

**Step-13 –** Make predictions on model.

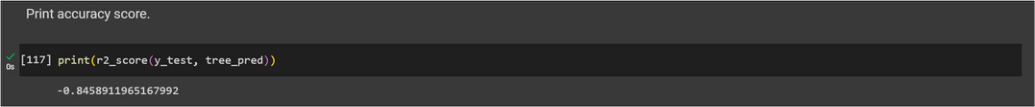
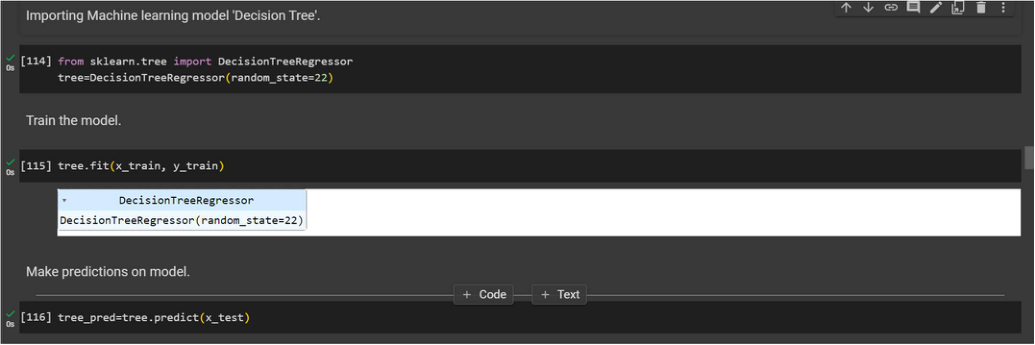
**Step-14 –** Checking accuracy score.

**Step-15 –** Import the Machine Learning Gradient Boost Model and train model

and then make prediction.

**Step-16 –** Check accuracy score of boosting algorithm.





SARVESH PRATAP YADAV

**Step-17 –** Import Decision Tree regressor Machine Learning Model and train   
 model and then make prediction.

**Step-18 –** Check accuracy score of Decision tree model.

**Conclusion –** Accurately predicting IMDb ratings of new movies is challenging. I needed   
to analyze combinations of features because no single feature can accurately predict   
ratings of new movies. Yet, there were numerous combinations of features that can affect   
IMDb ratings differently than the individual values would indicate. I performed a lot of data   
processing and two machine learning regression algorithms with Gradient Boosting   
technique. With Linear Regression algorithm, I got 2.80% of accuracy, then by using   
Gradient Boosting technique, I got 62% of accuracy and finally by using Decision Tree   
Regressor, I got 84% of accuracy. With this technique, I complete my machine learning   
model for prediction of movie rating with 84% of accuracy.

SARVESH PRATAP YADAV

2025