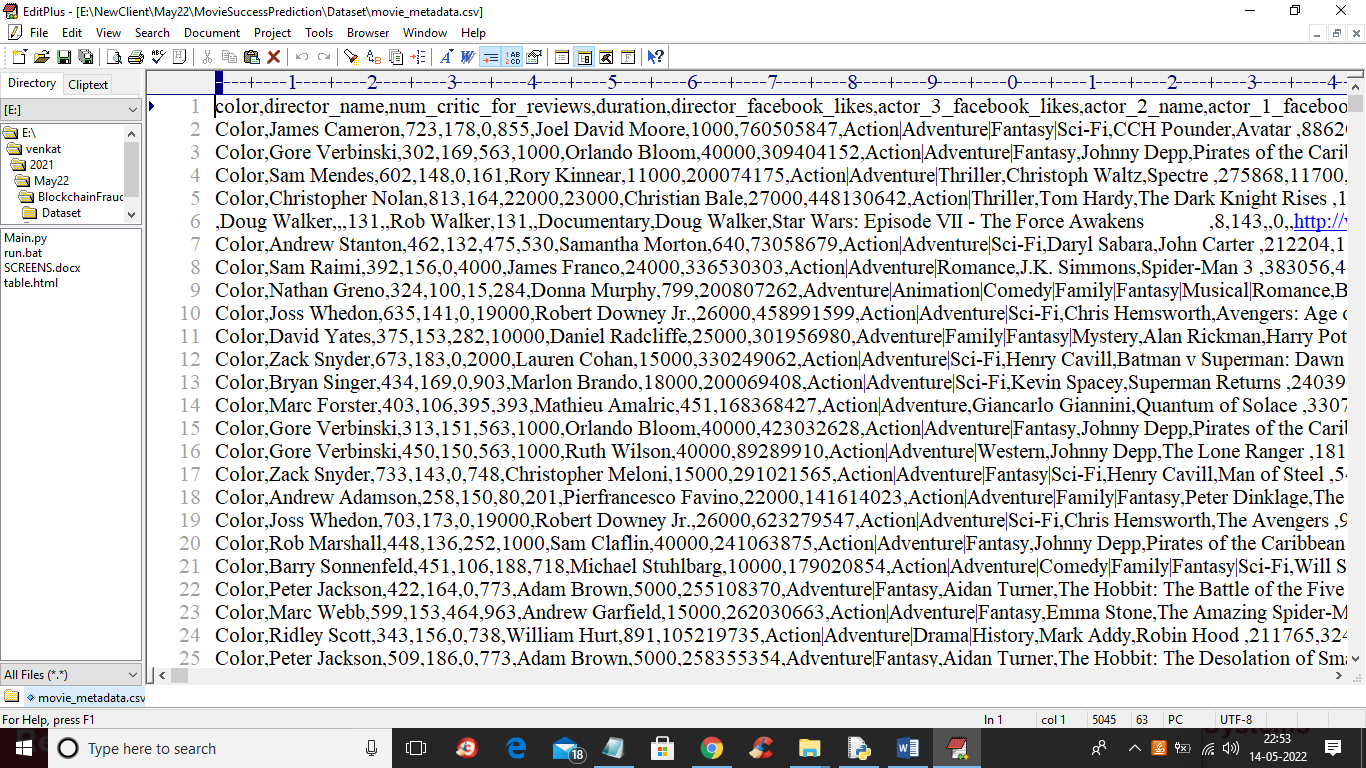
Movie Success Prediction Using Naïve Bayes, Logistic Regression and Support Vector Machine

The entertainment industry is a rapidly growing billion-dollar industry. With new milestones being reached almost every day, this industry has proved itself to be a very profitable business, if done correctly. Since huge investments are involved in the production and making of movies, both in terms of time and money, it would only make sense to try to predict the outcome beforehand. In an attempt to tackle this problem, we have built a model that predicts whether or not a movie can be called a success. The model compares the performance of three machine learning algorithms i.e. Naive Bayes, Logistic Regression, and Support Vector Machine (SVM), over two different datasets, to observe which performs better.

Experiment with real move dataset prove that logistic regression and SVM giving better accuracy.

To implement this project we have used MOVIE dataset from KAGGLE and below screen showing dataset details



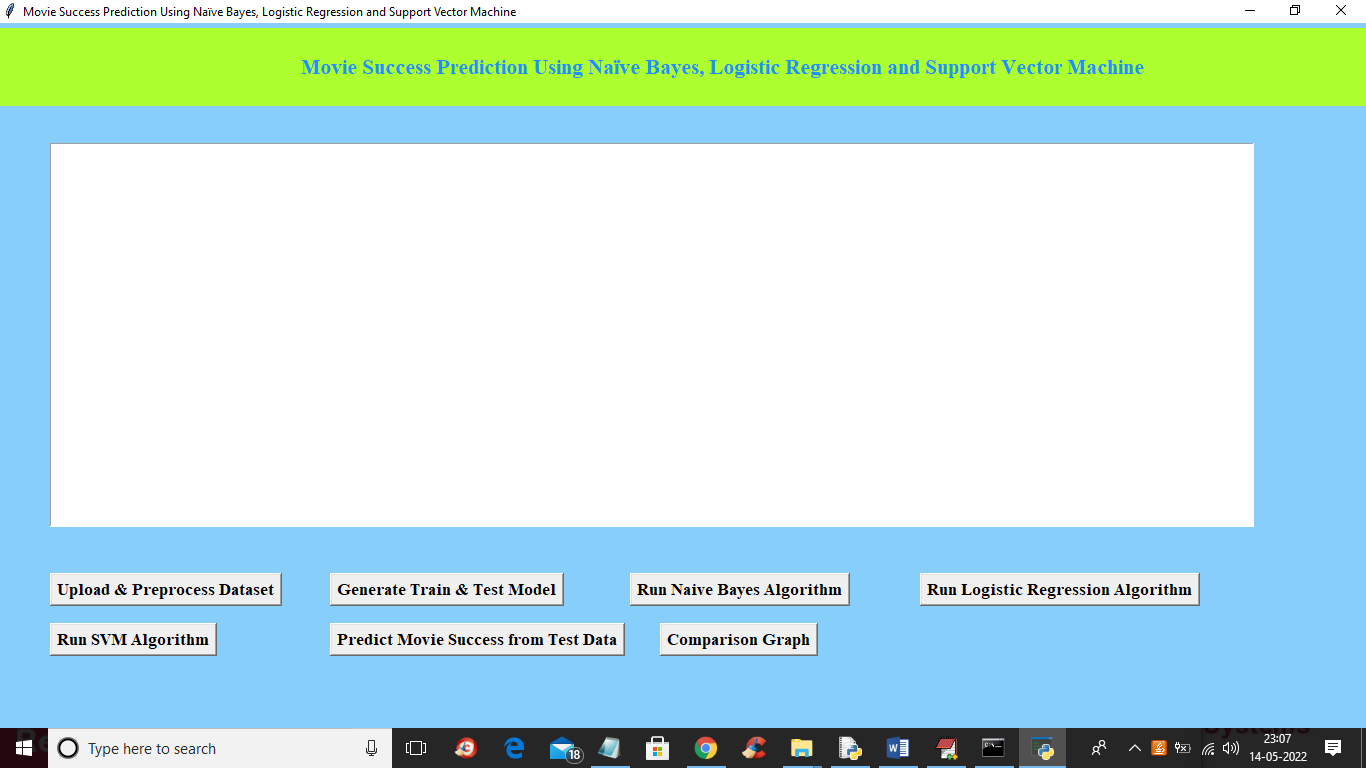
In above screen first row contains dataset column names and remaining rows are the dataset values and from above dataset we are using IMDB\_SCORE value to define movie category such as FLOP, AVG or HIT. If score 1 to 3 then movie is flop and 4 to 6 movie is AVG and greater than 6 will be consider as HIT.

To implement this project we have designed following modules

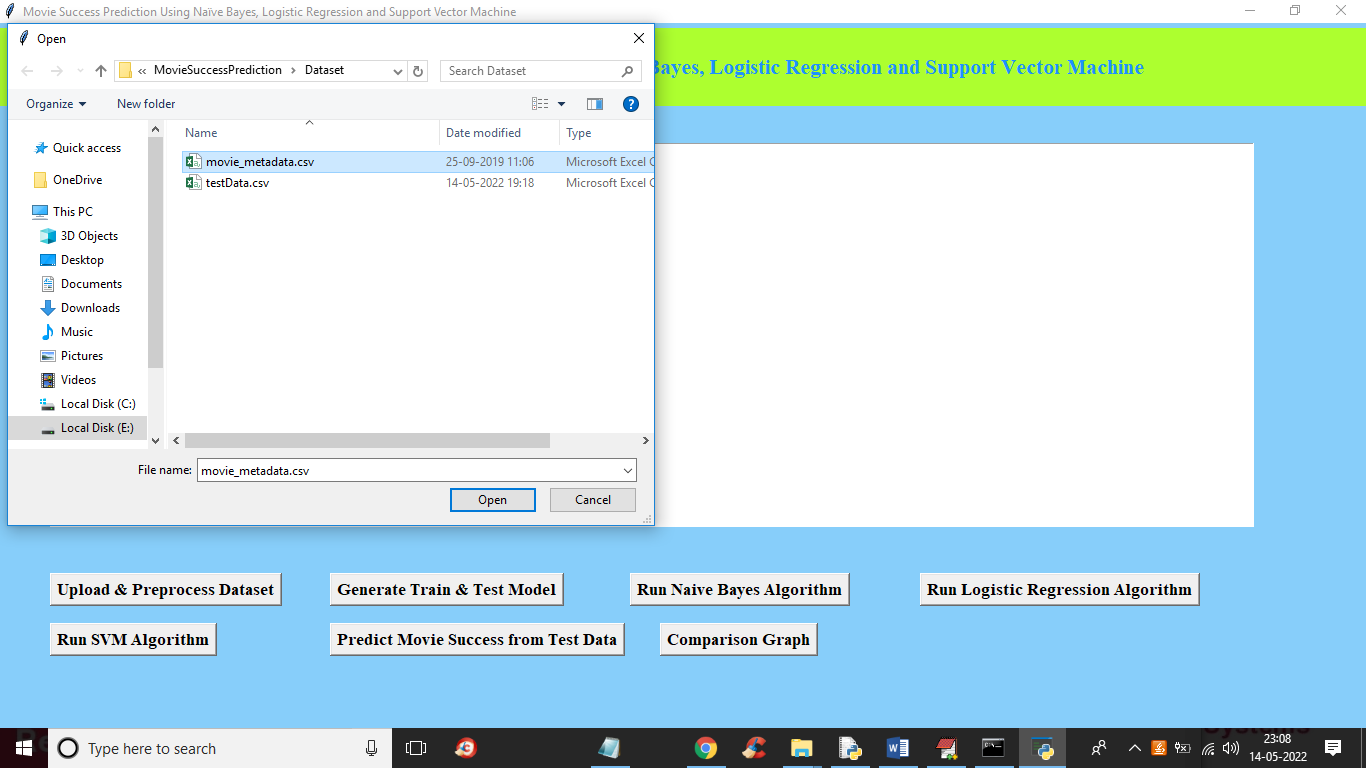
1. Upload & Preprocess Dataset: using this module we will upload MOVIE dataset and then remove missing values from dataset and then convert IMDB SCORE to movie category as FLOP, AVG or HIT
2. Generate Train & Test Model: using this module we will encode all non-numeric data into numeric data by using label encoder class and then split dataset into train and test
3. Run Naive Bayes Algorithm: using this module we will train Naïve Bayes algorithm by using train dataset and then apply trained Naïve Bayes algorithm on test data to calculate prediction accuracy
4. Run Logistic Regression Algorithm: using this module we will train Logistic Regression algorithm by using train dataset and then apply trained Logistic Regression algorithm on test data to calculate prediction accuracy
5. Run SVM Algorithm: using this module we will train SVM algorithm by using train dataset and then apply trained SVM algorithm on test data to calculate prediction accuracy
6. Predict Movie Success from Test Data: using this module we will upload test data and then Logistic Regression will predict movie category as HIT< FLOP or AVG.
7. Comparison Graph: using this module we will plot accuracy comparison graph between all algorithms

SCREEN SHOTS

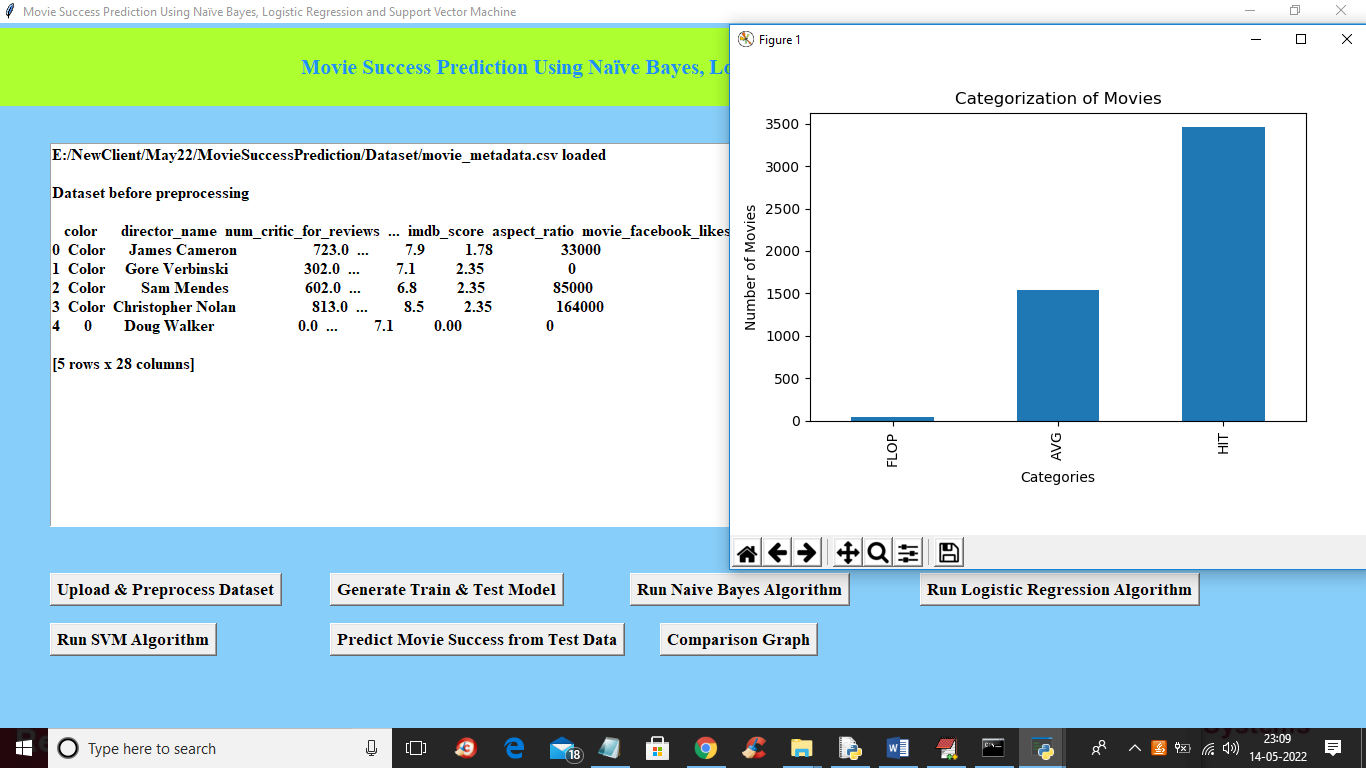
To run project double click on ‘run.bat’ file to get below screen



In above screen click on ‘Upload & Preprocess Dataset’ button to load dataset and to get below screen



In above screen selecting and uploading movie dataset and then click on ‘Open’ button to load dataset and then get below output



In above screen dataset loaded and in graph x-axis contains movie category and y-axis represents count of that movie category and in above screen we can see dataset contains some non-numeric data so close above graph and then click on ‘Generate Train & Test Model’ button to encode non-numeric data to numeric data and then split dataset into train and test. All ML algorithms will take only numeric dataset so we need to convert all non-numeric data to numeric and get below output



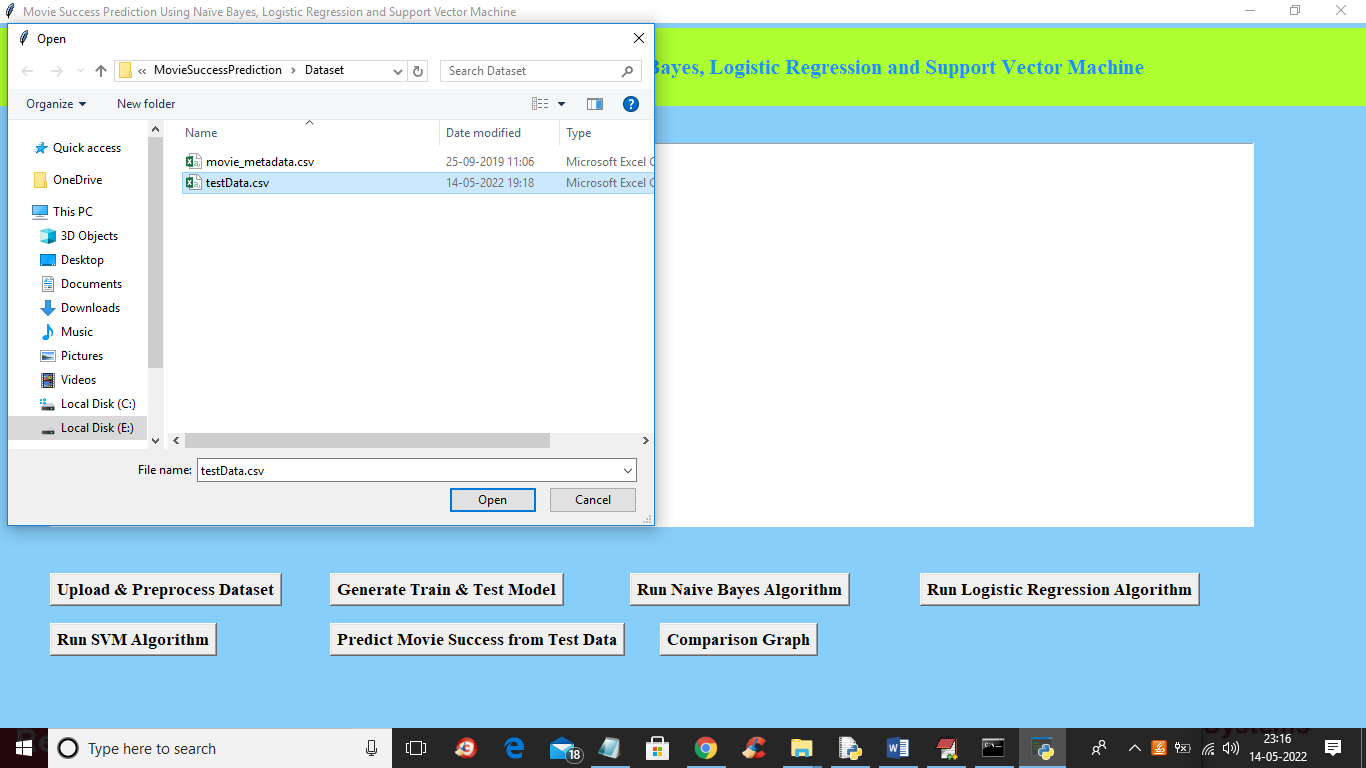
In above screen we can see all dataset values converted to numeric and then we can see dataset contains 5043 records and 23 columns and then we split 80% dataset for training and 20% dataset for testing and now train and test data is ready and now click on ‘Run Naïve Bayes Algorithm’ button to get below output



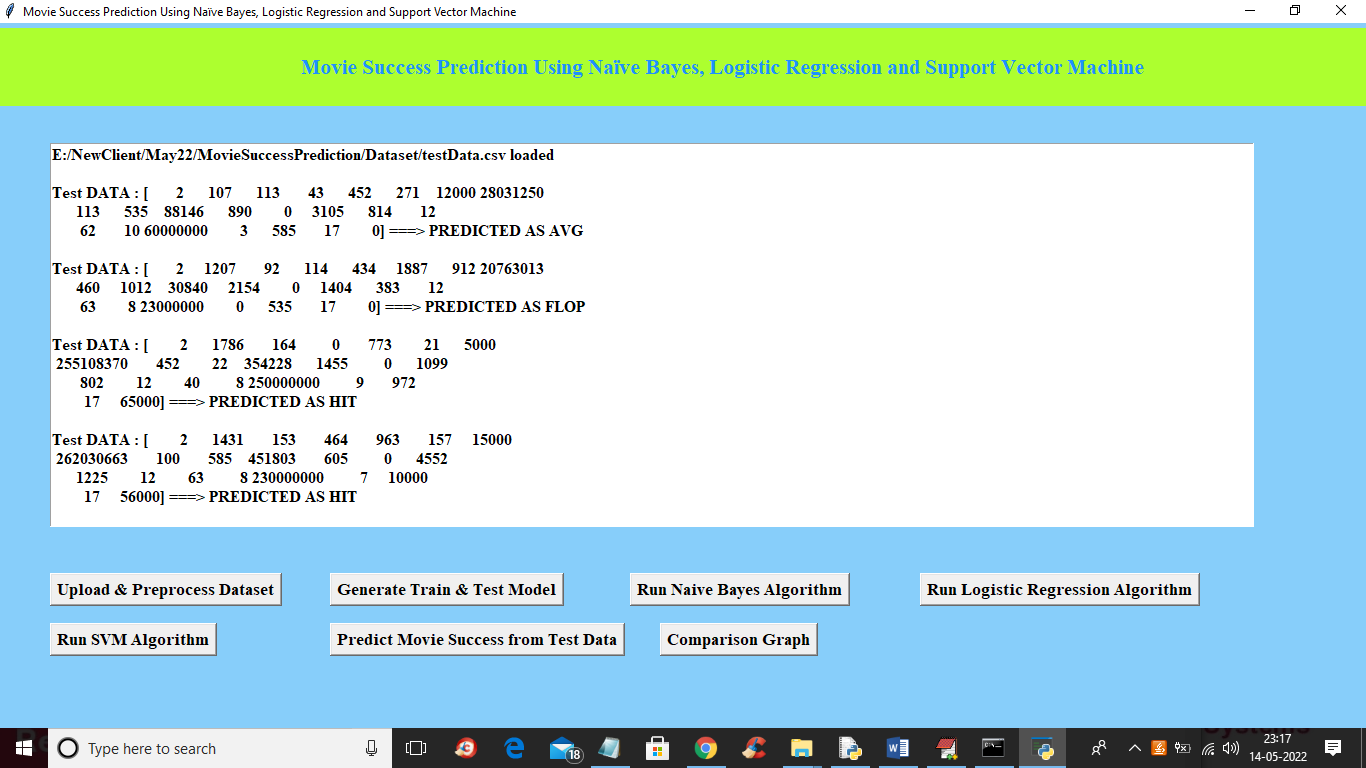
In above screen with Naïve Bayes we got 35% accuracy and now run Logistic Regression and SVM algorithm



In above screen with Logistic Regression we got 99% accuracy and with SVM we got 79% accuracy and now click on ‘Predict Movie Success from Test Data’ button to upload test data and predict movie success



In above screen selecting and uploading ‘testData.csv’ file and then click on ‘Open’ button to get below output



In above screen in square bracket we can see test data and then after square bracket ==🡺 arrow symbol we got predicted movie success result as HIT, AVG or FLOP