ROTTEN TOMATOES GENRE PREDICTION MODEL

Goal : The project aims to predict various movie genres.

Process:

1. The dataset was scrapped from Rotten Tomatoes website using the selenium web driver.

2. Data was cleaned and filtered in Excel to keep only the columns that were useful for the project.

3. Applied multiple machine learning algorithms to the dataset to predict the genres for movies.

Outputs:

1) Naïve Bayes

```
1   Naive Bayes Output:
2
3   for genre sci-fi , f1 score is 0.71
4   for genre horror , f1 score is 0.70
5   for genre fantasy , f1 score is 0.66
6   for genre adventure , f1 score is 0.67
7   for genre thriller , f1 score is 0.56
8   for genre mystery , f1 score is 0.64
9   for genre romance , f1 score is 0.61
10  for genre crime , f1 score is 0.62
11  for genre drama , f1 score is 0.50
12  for genre action , f1 score is 0.64
13  for genre comedy , f1 score is 0.58
14  for genre documentary , f1 score is 0.67
15  for genre war , f1 score is 0.65
16  average f1 score : 0.63
```

2) SVM

```
SVM Output:

for "sci-fi" , f1 score = 0.70
for "horror" , f1 score = 0.68
for "fantasy" , f1 score = 0.62
for "adventure" , f1 score = 0.66
for "thriller" , f1 score = 0.63
for "mystery" , f1 score = 0.58
for "romance" , f1 score = 0.62
for "crime" , f1 score = 0.56
for "drama" , f1 score = 0.59
for "action" , f1 score = 0.67
for "comedy" , f1 score = 0.62
for "documentary" , f1 score = 0.64
for "war" , f1 score = 0.65
average f1 score over all genres : 0.63
```

3) Logistic Regression

```
Logistic Regression Output:

for "sci-fi" , f1 score = 0.66
for "horror" , f1 score = 0.68
for "fantasy" , f1 score = 0.62
for "adventure" , f1 score = 0.65
for "thriller" , f1 score = 0.62
for "mystery" , f1 score = 0.57
for "romance" , f1 score = 0.60
for "crime" , f1 score = 0.59
for "drama" , f1 score = 0.57
for "action" , f1 score = 0.66
for "comedy" , f1 score = 0.61
for "documentary" , f1 score = 0.62
for "war" , f1 score = 0.66
average f1 score over all genres : 0.60
```

4)K-Nearest Neighbor

```
KNN Output:

for "sci-fi" , f1 score = 0.65
for "horror" , f1 score = 0.68
for "fantasy" , f1 score = 0.63
for "adventure" , f1 score = 0.64
for "thriller" , f1 score = 0.52
for "mystery" , f1 score = 0.55
for "romance" , f1 score = 0.54
for "crime" , f1 score = 0.48
for "drama" , f1 score = 0.48
for "action" , f1 score = 0.59
for "comedy" , f1 score = 0.56
for "documentary" , f1 score = 0.63
for "war" , f1 score = 0.59
average f1 score over all genres : 0.61
```

5) Random Forest

```
Random Forest Output:

for "sci-fi" , f1 score = 0.63
for "horror" , f1 score = 0.64
for "fantasy" , f1 score = 0.61
for "adventure" , f1 score = 0.60
for "thriller" , f1 score = 0.58
for "mystery" , f1 score = 0.53
for "romance" , f1 score = 0.54
for "crime" , f1 score = 0.57
for "drama" , f1 score = 0.54
for "action" , f1 score = 0.64
for "comedy" , f1 score = 0.58
for "documentary" , f1 score = 0.59
for "war" , f1 score = 0.65
average f1 score over all genres : 0.59
```

Conclusion:

The project was successfully completed, and all the Machine Learning Algorithms worked perfectly fine with False negatives less than 70%.

The best Algorithms in predicting the genres for the movies are Naive Bayes, SVM and Logistic Regression.

On the other hand, Random Forest worked okay but gave an average score of 0.59 which was less than the other Algorithms.