Steps to run the Project:

1). Run project on Anaconda Navigator (JupyterLab) – Version: 1.9.7

2). Install and import the libraries if not available

- import string
- import pandas as pd
- import numpy as np
- import Spacy
- from nltk.corpus import stopwords
- from sklearn.feature_extraction.stop_words import ENGLISH_STOP_WORDS
- from sklearn.feature_extraction.text import TfidfVectorizer, CountVectorizer
- from sklearn.linear_model import LogisticRegression
- from sklearn.metrics import f1_score, accuracy_score, r2_score
- from sklearn.model_selection import (train_test_split, learning_curve, cross_val_score, cross_val_predict,
- ShuffleSplit, KFold)
- import time
- from sklearn.ensemble import RandomForestClassifier
- from sklearn.naive_bayes import MultinomialNB
- from sklearn import metrics
- from sklearn import svm
- from sklearn.svm import SVC
- import matplotlib.pyplot as plt
- from sklearn.model_selection import learning_curve
- from sklearn.model_selection import ShuffleSplit
- from sklearn.preprocessing import LabelEncoder

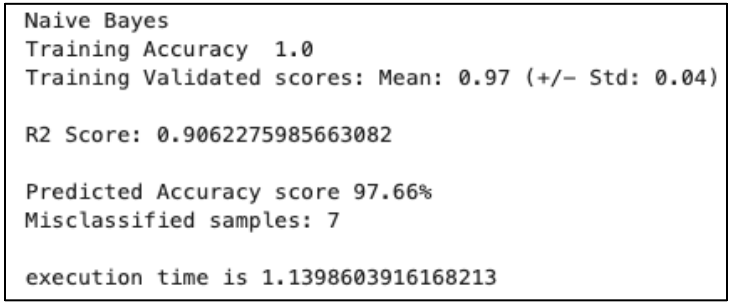
3). We have used Anaconda (JupyterLab): files included in archive.

- BiGram Classifier – Results based on bigram words of 1000 records.
- Latest Data Analysis – Results based on unigram words of 1000 records.
- Multi Record Classifier – Results of 3000 records.
- CombineCSV – Combine true and false csv’s.
- Sample Dataset – contains 1000 records of true and false records(csv).
- SampleDataset500, SampleDataset1000, SampleDataset1500(csv).

Output Explanation:

Results of Classifiers used in this project:

1). Naive Bayes:



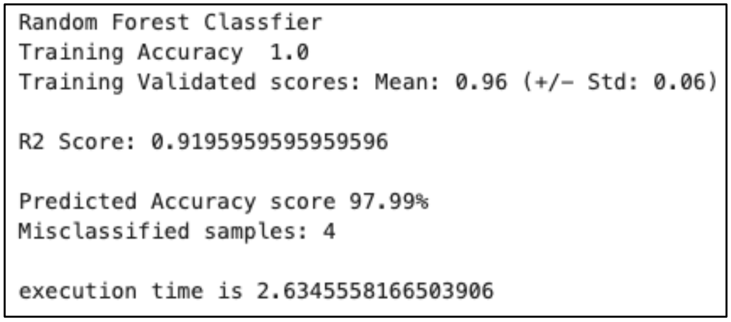
```
Naive Bayes
Training Accuracy  1.0
Training Validated scores: Mean: 0.97 (+/- Std: 0.04)

R2 Score: 0.9062275985663082

Predicted Accuracy score 97.66%
Misclassified samples: 7

execution time is 1.1398603916168213
```

2). Random Forest:



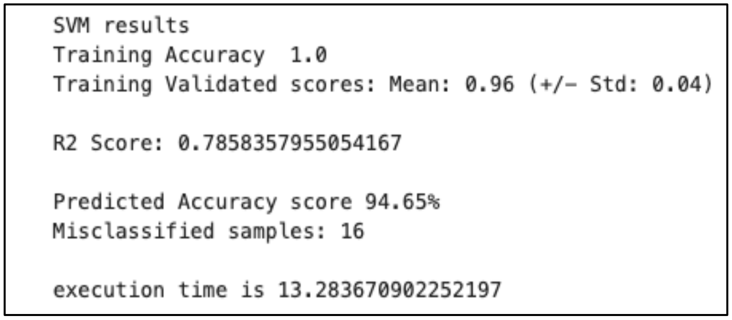
```
Random Forest Classfier
Training Accuracy  1.0
Training Validated scores: Mean: 0.96 (+/- Std: 0.06)

R2 Score: 0.9195959595959596

Predicted Accuracy score 97.99%
Misclassified samples: 4

execution time is 2.6345558166503906
```

3). SVM:



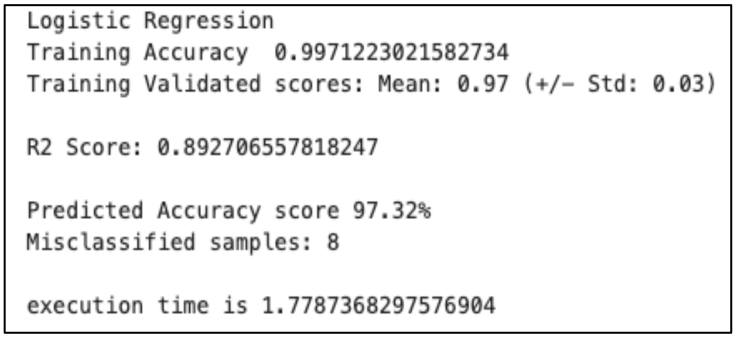
```
SVM results
Training Accuracy  1.0
Training Validated scores: Mean: 0.96 (+/- Std: 0.04)

R2 Score: 0.7858357955054167

Predicted Accuracy score 94.65%
Misclassified samples: 16

execution time is 13.283670902252197
```

4). Logistic Regression:



```
Logistic Regression
Training Accuracy  0.9971223021582734
Training Validated scores: Mean: 0.97 (+/- Std: 0.03)

R2 Score: 0.892706557818247

Predicted Accuracy score 97.32%
Misclassified samples: 8

execution time is 1.7787368297576904
```