

# REPORT ON

## SPAM NEWS DETECTION

With the rise of digital media and social networking, the distribution of fake or misleading news articles, often referred to as "spam news," has become a significant issue. Such misinformation can have adverse effects on public opinion, political discourse, and much more. This project focuses on developing a system for the automated detection of spam news articles using machine learning techniques.

### **Objective:**

The main objective of this project is to design and implement a spam news detection system capable of accurately identifying and filtering out fake or misleading news articles from legitimate sources.

This Project is coded in python programming where the program is programmed in 5 sectors they are:-

1. **Loading the Dataset:-**

The given set of data with features are loaded in the form of either Mammogram Images or CSV File and here two datasets named True news and Fake news have been loaded.

2. **Cleaning the loaded data:-**

The collected data undergo preprocessing techniques such as resizing, normalization, and noise removal to enhance their quality and suitability for analysis.

Features such as word frequency, sentiment analysis, and readability scores are extracted from the pre-processed text using natural language processing (NLP) techniques.

3. **Conversion of Categorical Data:-**

The conversion of categorical data involves transforming textual features extracted from news articles into numerical representations that can be used by machine learning algorithms. This conversion process facilitates the analysis and classification of news articles by converting categorical variables such as word presence or absence, sentiment scores, or readability metrics into numerical values, enabling the models to effectively learn and distinguish between spam and legitimate news content.

4. **Fitting the training Data through the relevant Training Model:-**

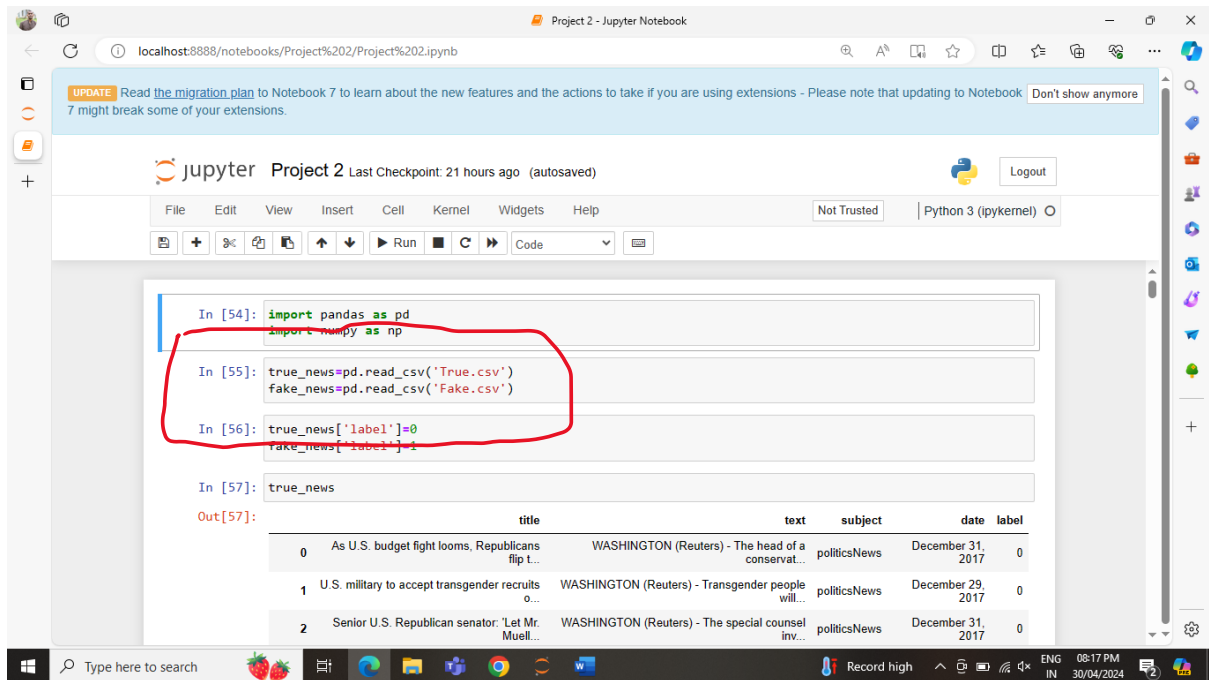
The trained Model is Fitted into the relevant training Models like KNN, NAÏVE BAYES, LOGISTIC REGRESSION and much more

5. **Prediction:-**

The performance of the developed models is evaluated using metrics like accuracy, sensitivity, specificity, and area under the receiver operating characteristic curve (AUC-ROC) through cross-validation techniques and Once the model achieves satisfactory performance, it is deployed as a user-friendly application or integrated into existing healthcare systems for real-time breast cancer detection.

Some of the screenshots of the Model is displayed below:-

## 1. Loading the Dataset:-



The screenshot shows a Jupyter Notebook interface with the following code cells:

```
In [54]: import pandas as pd
import numpy as np

In [55]: true_news=pd.read_csv('True.csv')
fake_news=pd.read_csv('Fake.csv')
```

The cells for In [55] and In [56] are circled in red. The output of In [57] is displayed as a table:

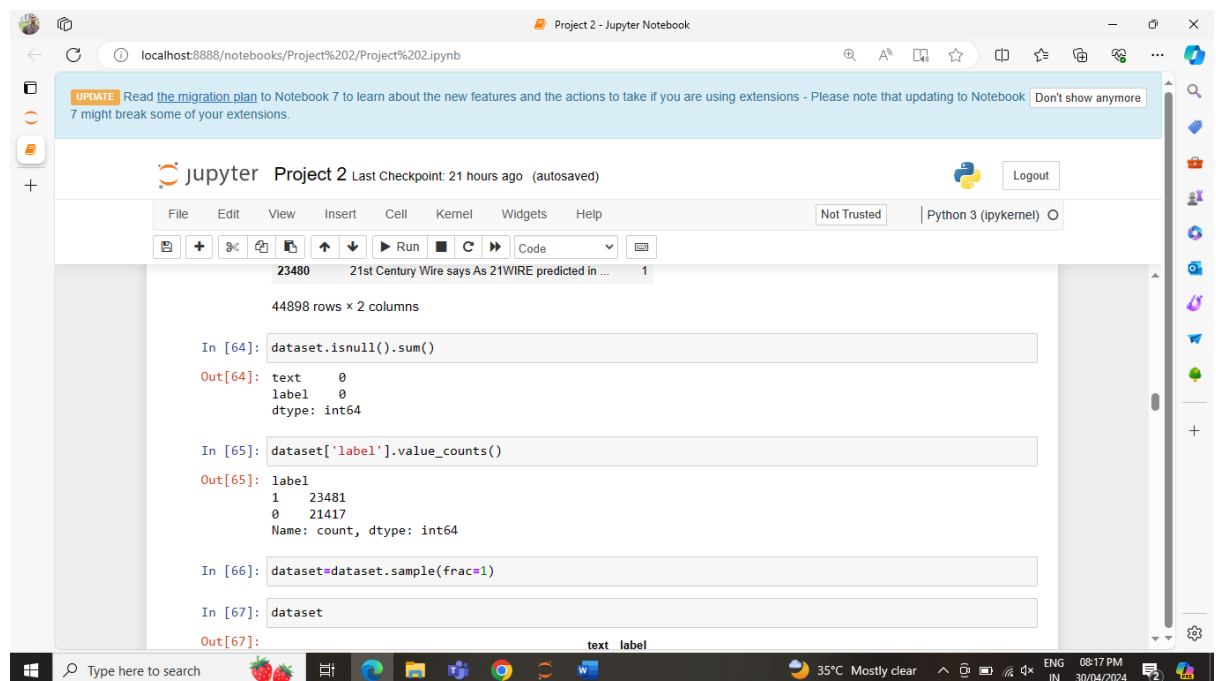
```
In [56]: true_news['label']=0
fake_news['label']=1

In [57]: true_news

Out[57]:
```

	title	text	subject	date	label
0	As U.S. budget fight looms, Republicans flip t...	WASHINGTON (Reuters) - The head of a conservat...	politicsNews	December 31, 2017	0
1	U.S. military to accept transgender recruits o...	WASHINGTON (Reuters) - Transgender people will...	politicsNews	December 29, 2017	0
2	Senior U.S. Republican senator: 'Let Mr. Muell...	WASHINGTON (Reuters) - The special counsel inv...	politicsNews	December 31, 2017	0

## 2. Cleaning the loaded data:-



The screenshot shows a Jupyter Notebook interface with the following code cells:

```
23480 21st Century Wire says As 21WIRE predicted in ... 1

44898 rows x 2 columns

In [64]: dataset.isnull().sum()

Out[64]:
text      0
label     0
dtype: int64

In [65]: dataset['label'].value_counts()

Out[65]:
label
1    23481
0    21417
Name: count, dtype: int64

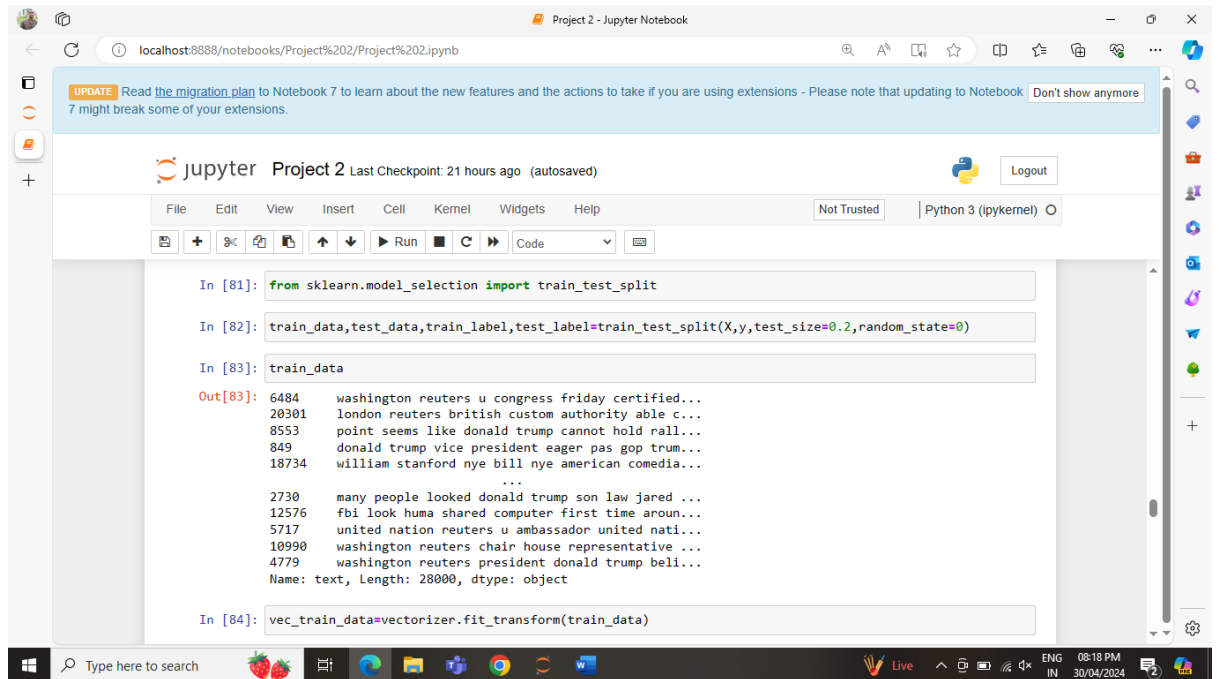
In [66]: dataset=dataset.sample(frac=1)

In [67]: dataset

Out[67]:
```

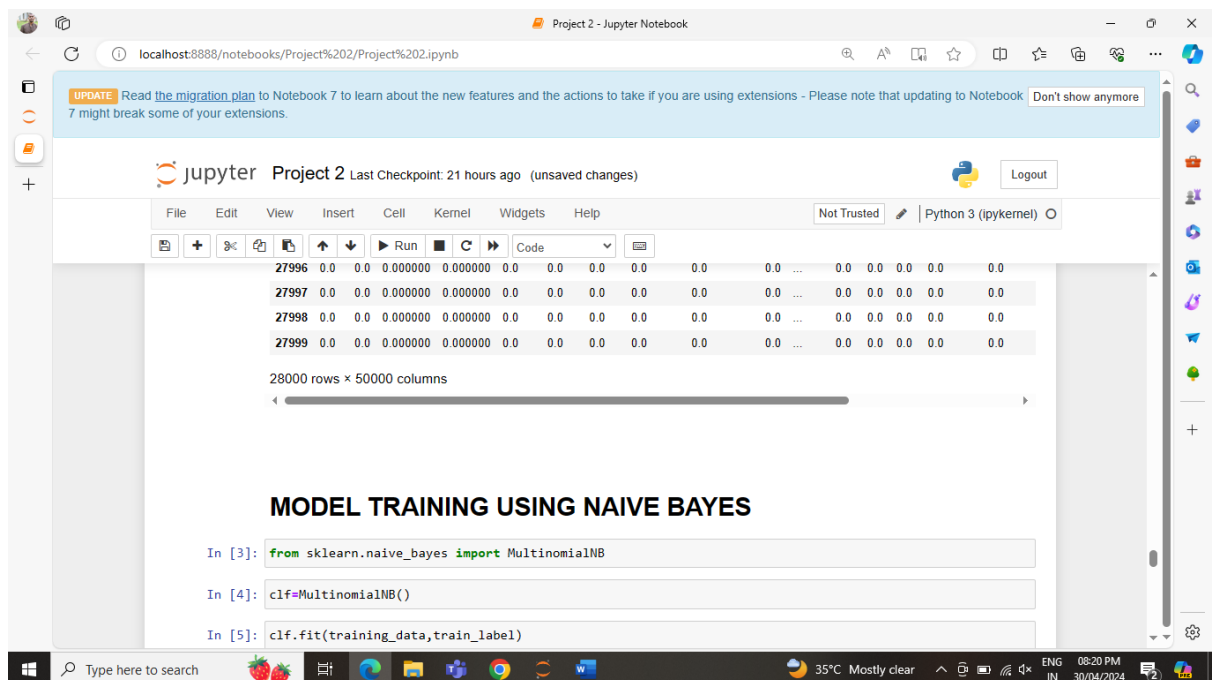
The output of In [67] is a table with columns 'text' and 'label'.

### 3. Conversion of Categorical Data:-

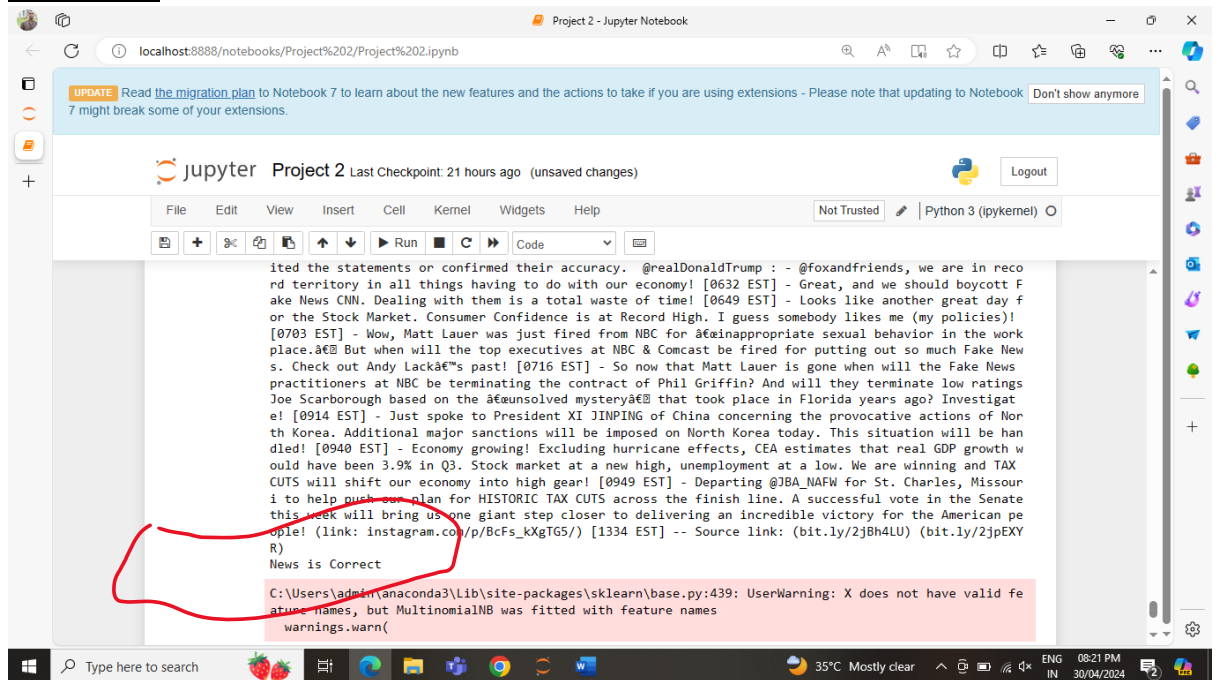


#### 4. Fitting the training Data through the relevant Training Model:-

Modelling is done through Naïve Bayes Model as it is efficient in giving the accuracy for the trained and tested data



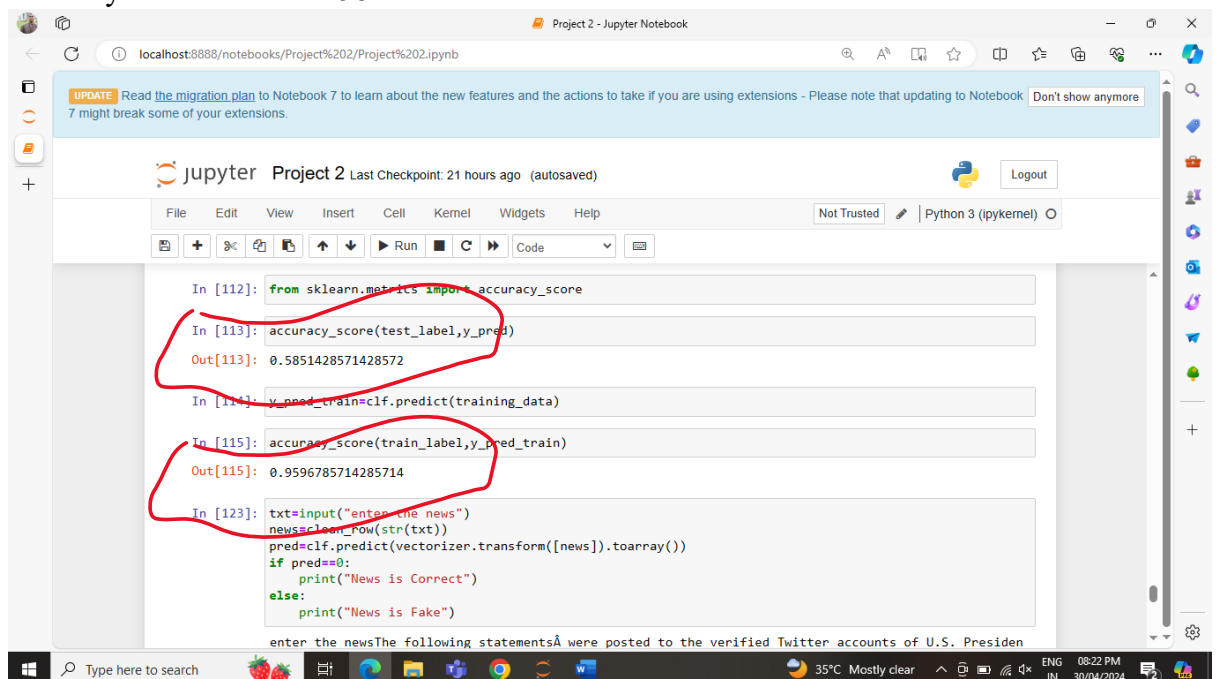
## 5. Prediction:-



```
ited the statements or confirmed their accuracy. @realDonaldTrump : - @foxandfriends, we are in record territory in all things having to do with our economy! [0632 EST] - Great, and we should boycott Fake News CNN. Dealing with them is a total waste of time! [0649 EST] - Looks like another great day for the Stock Market. Consumer Confidence is at Record High. I guess somebody likes me (my policies)! [0703 EST] - Wow, Matt Lauer was just fired from NBC for inappropriate sexual behavior in the workplace. But when will the top executives at NBC & Comcast be fired for putting out so much Fake News. Check out Andy Lack's past! [0716 EST] - So now that Matt Lauer is gone when will the Fake News practitioners at NBC be terminating the contract of Phil Griffin? And will they terminate low ratings Joe Scarborough based on the unsolved mystery that took place in Florida years ago? Investigate! [0914 EST] - Just spoke to President XI JINPING of China concerning the provocative actions of North Korea. Additional major sanctions will be imposed on North Korea today. This situation will be handled! [0940 EST] - Economy growing! Excluding hurricane effects, CEA estimates that real GDP growth would have been 3.9% in Q3. Stock market at a new high, unemployment at a low. We are winning and TAX CUTS will shift our economy into high gear! [0949 EST] - Departing @JBA_NAFW for St. Charles, Missouri to help push our plan for HISTORIC TAX CUTS across the finish line. A successful vote in the Senate this week will bring us one giant step closer to delivering an incredible victory for the American people! (link: instagram.com/p/BcFs_kXgTG5/) [1334 EST] -- Source link: (bit.ly/2jBh4LU) (bit.ly/2jpEXYR)
News is Correct

C:\Users\admin\anaconda3\Lib\site-packages\sklearn\base.py:439: UserWarning: X does not have valid feature names, but MultinomialNB was fitted with feature names
warnings.warn(
```

As per the modelling Naïve bayes Modelling the accuracy of trained data is 95% and accuracy of tested data is 58%



```
In [112]: from sklearn.metrics import accuracy_score
In [113]: accuracy_score(test_label,y_pred)
Out[113]: 0.5851428571428572

In [114]: y_pred_train=clf.predict(training_data)
In [115]: accuracy_score(train_label,y_pred_train)
Out[115]: 0.9596785714285714

In [123]: txt=input("enter the news")
news=token_row(str(txt))
pred=clf.predict(vectorizer.transform([news]).toarray())
if pred==0:
    print("News is Correct")
else:
    print("News is Fake")

enter the newsThe following statements were posted to the verified Twitter accounts of U.S. Presiden
```

**Conclusion:-**

The project successfully demonstrates the feasibility and effectiveness of using machine learning techniques for automated spam news detection. The developed system can play a crucial role in combating the spread of misinformation online and promoting media literacy among internet users.