

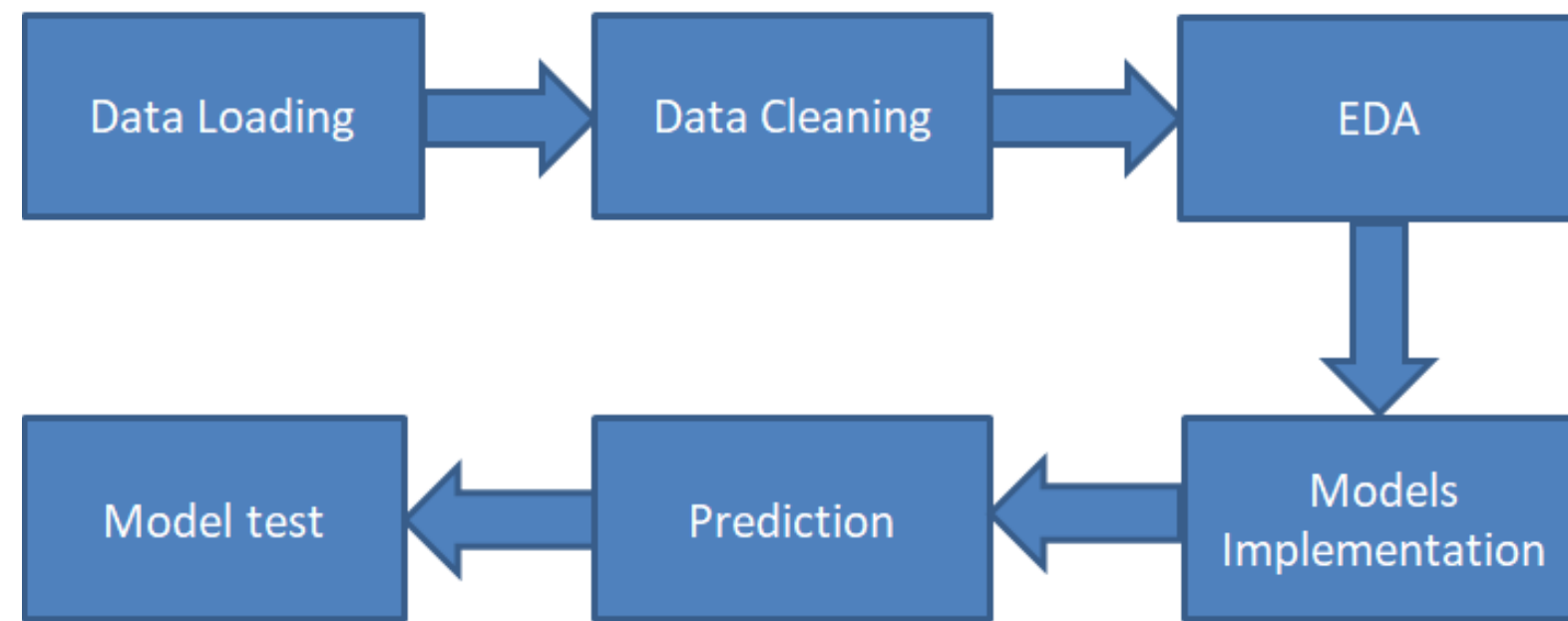
Made by: Abdelaziz

IRON  
HACK

FAKKE  
NEWS!

DAFT NOV21

1. Fake news becomes to be a significant issue on the internet and social media affecting the mental and physical health of children and adults.
2. Once a source is labeled as a producer of fake news, we can predict with high level of confidence that any future articles from that source will also be fake news.





# About data

Columns	Isna().sum()
id	0
title	558
author	1957
text	39
label	0

	Columns	Count	Non_Null	Dtype
0	id	20800	Non_null	int64
1	title	20800	Non_null	object
2	author	20800	Non_null	object
3	text	20800	Non_null	object
4	label	20800	Non_null	Int64

NewsF.shape()

20800

5

NewsF.duplicated().sum()

0

# Data Cleaning/Processing

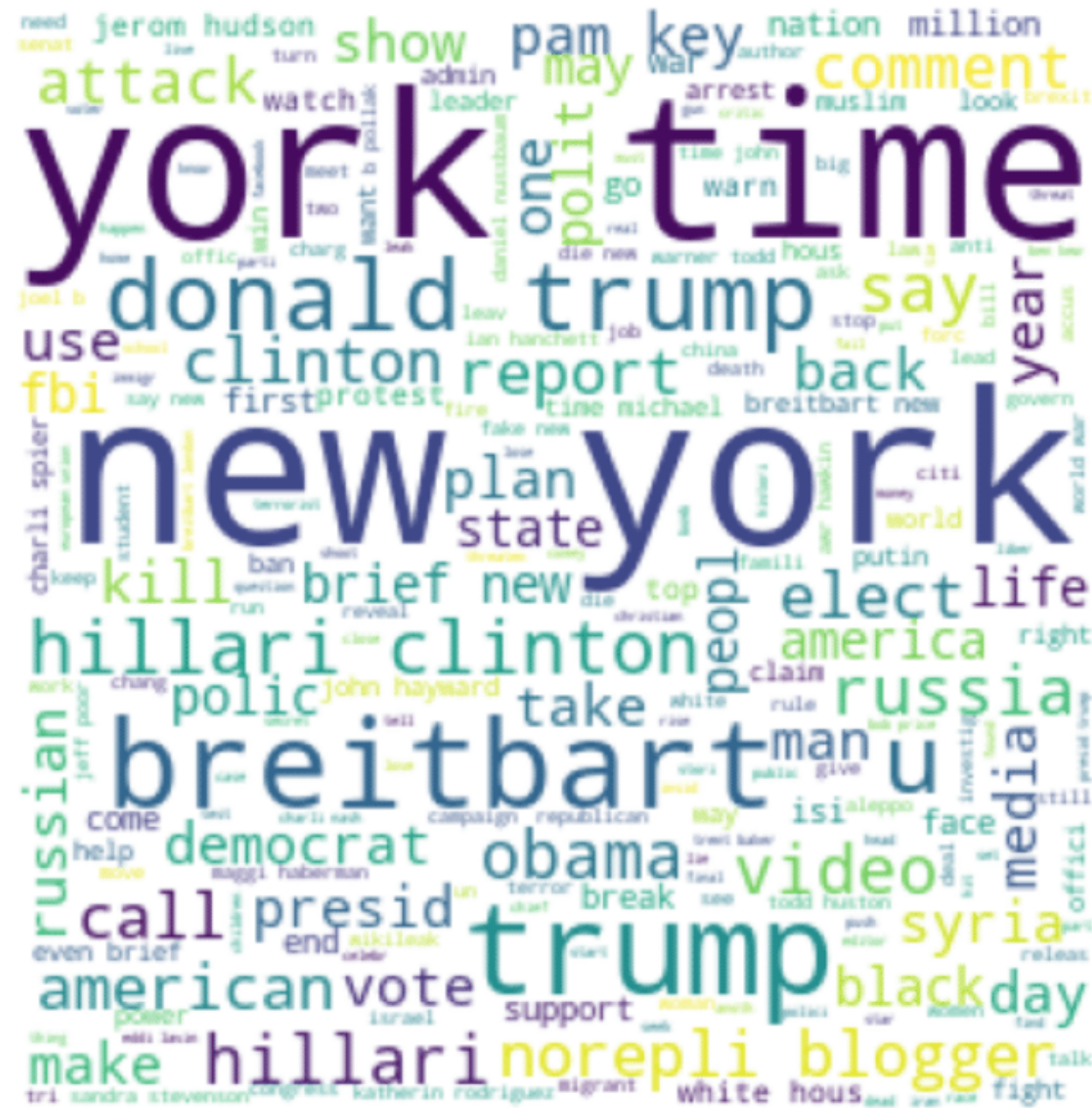
```
# we can remove all null values from ower NewsF dataset  
NewsF = NewsF.fillna('')
```

```
NewsF['Merge'] = NewsF['author'] + " " + NewsF['title']
```

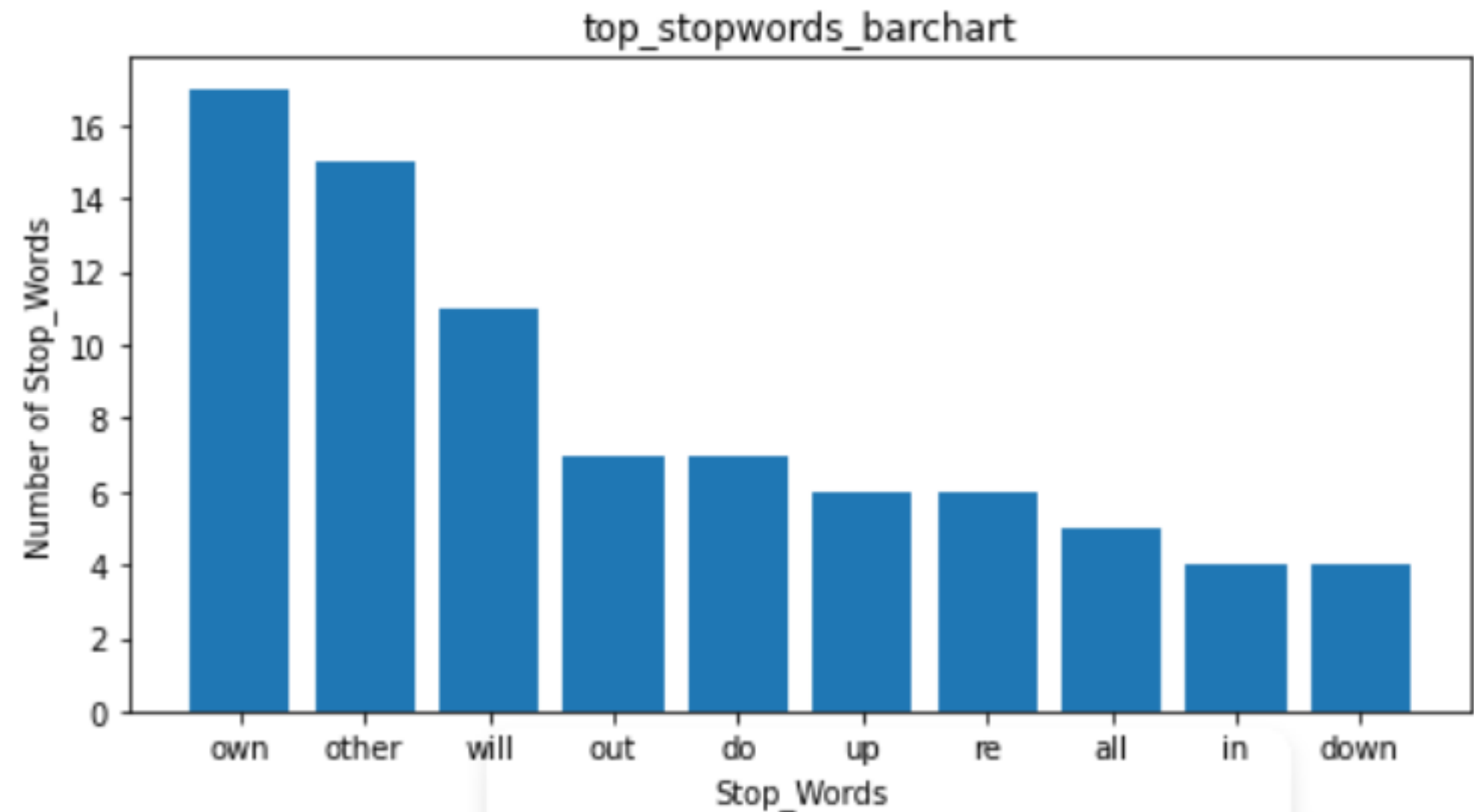
```
stem = PorterStemmer()  
def stemming(Merge):  
    stemmed_Merge = re.sub('[^a-zA-Z]', ' ', Merge)  
    stemmed_Merge = stemmed_Merge.split()  
    stemmed_Merge = [stem.stem(word) for word in stemmed_Merge if not word in stopwords.words('english')]  
    stemmed_Merge = ' '.join(stemmed_Merge)  
    return stemmed_Merge
```

```
NewsF['Merge'] = NewsF['Merge'].apply(stemming)
```

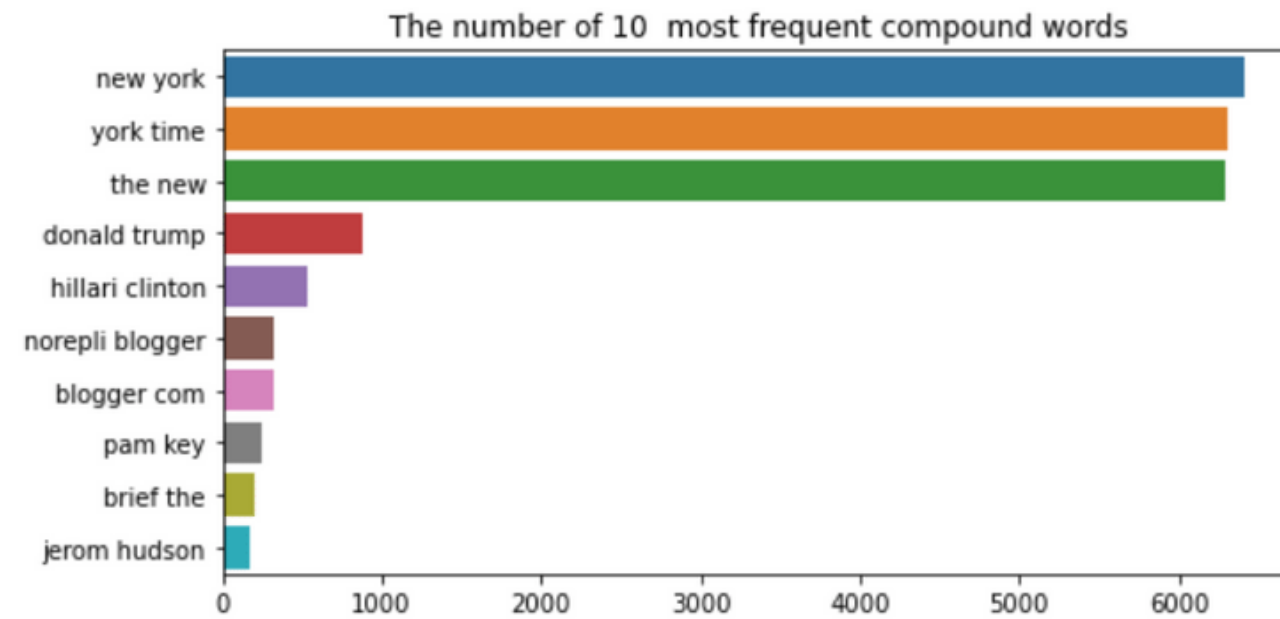
# Word Cloud



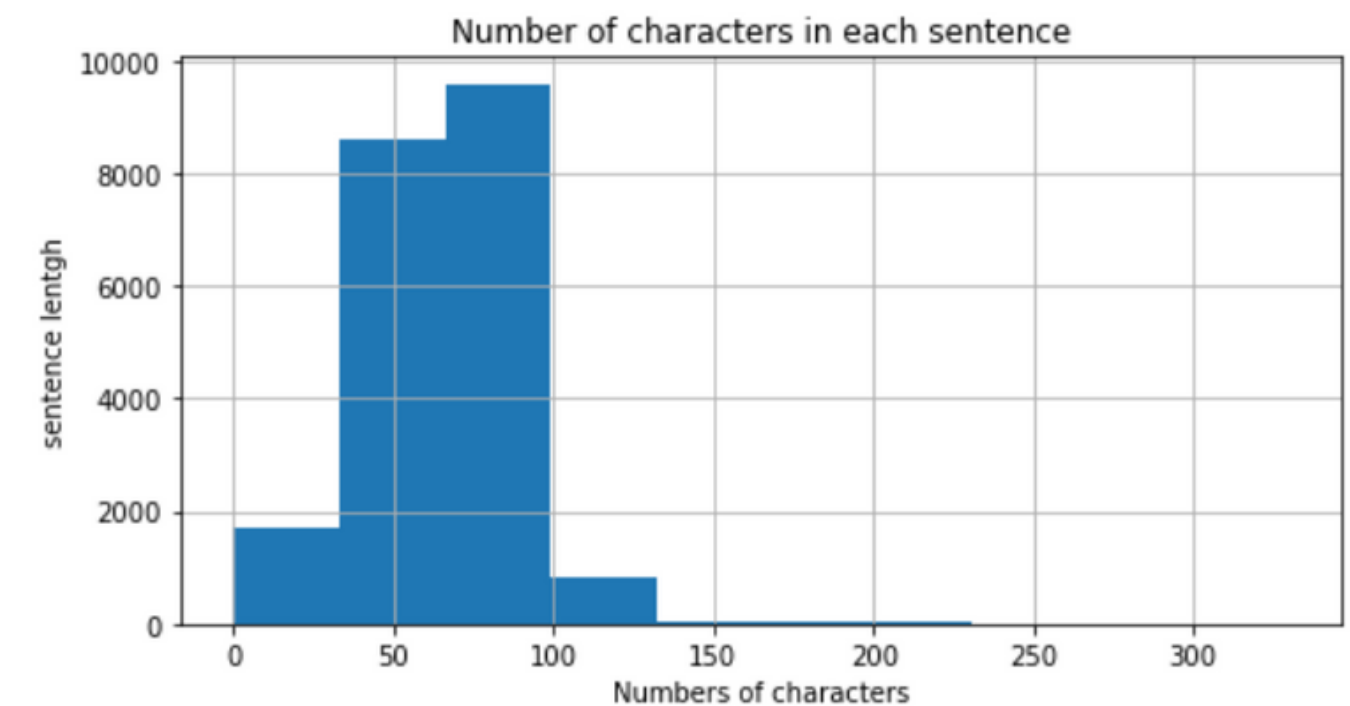
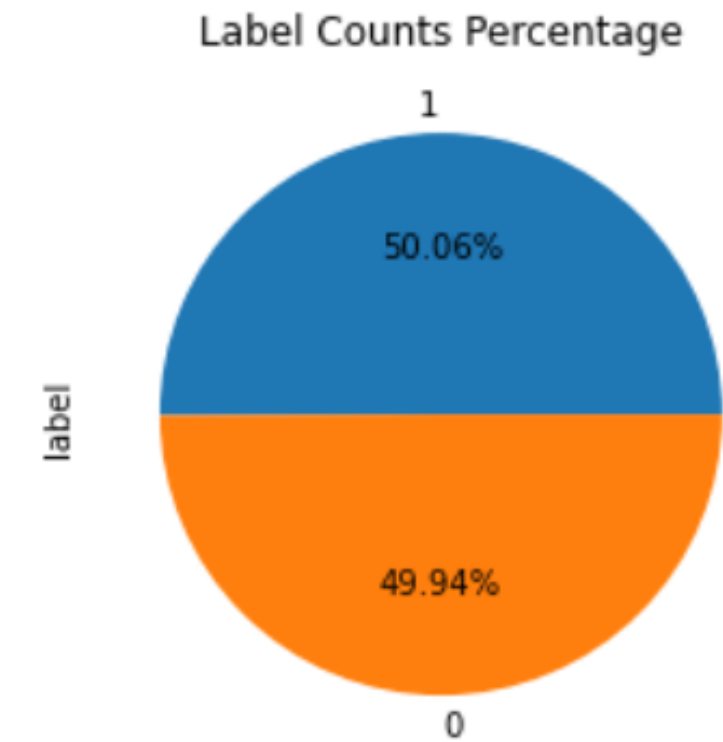
# EDA



# EDA



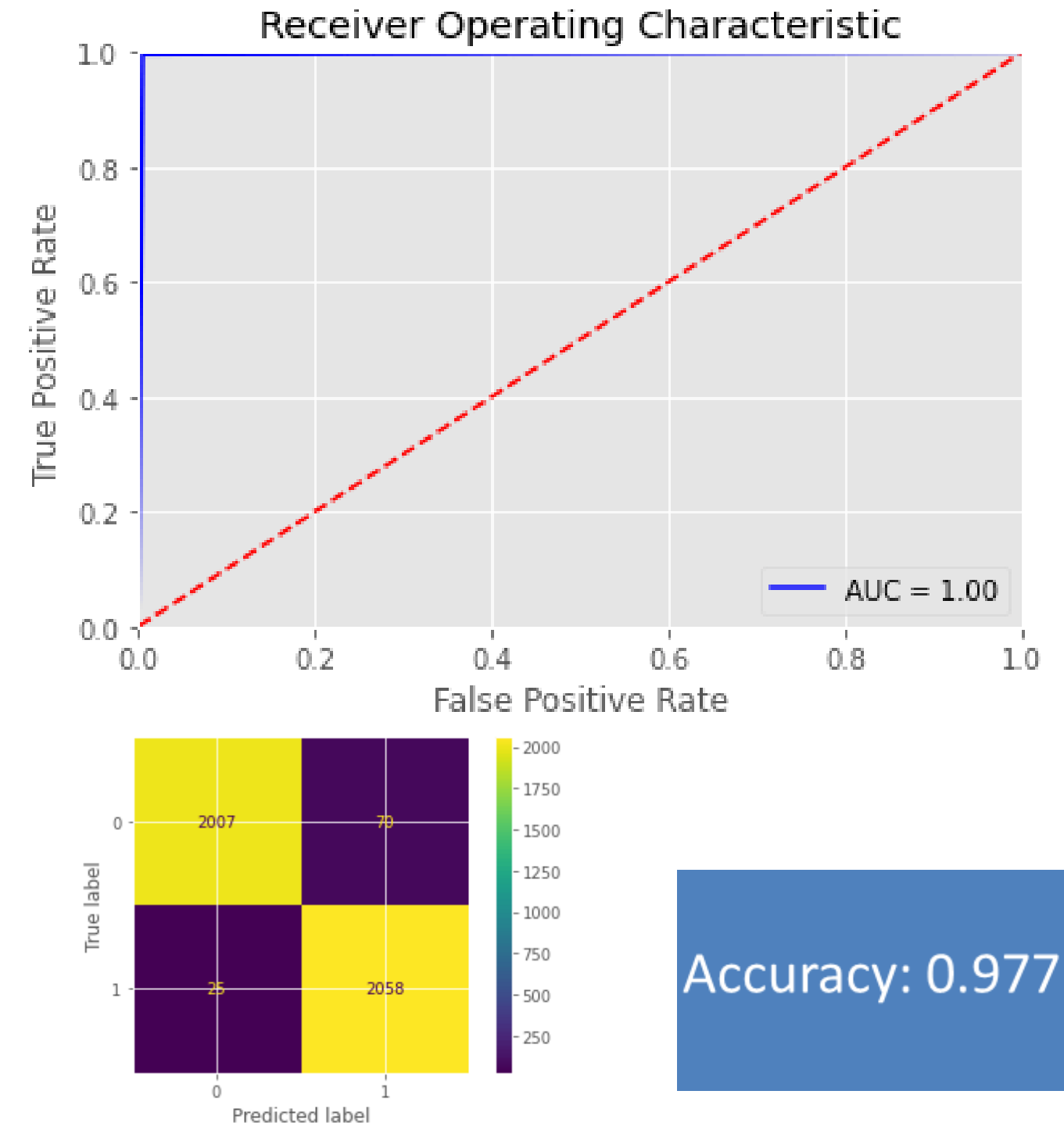
	Words	Total Count	Percentage
0	new	7,167	3.3%
1	time	6,675	3.1%
2	york	6,415	2.9%
3	trump	3,740	1.7%
4	breitbart	2,756	1.3%
5	clinton	1,272	0.6%
6	hillari	1,267	0.6%
7	donald	905	0.4%
8	news	769	0.4%
9	elect	768	0.4%



# Supervised models

## Logistic Regression

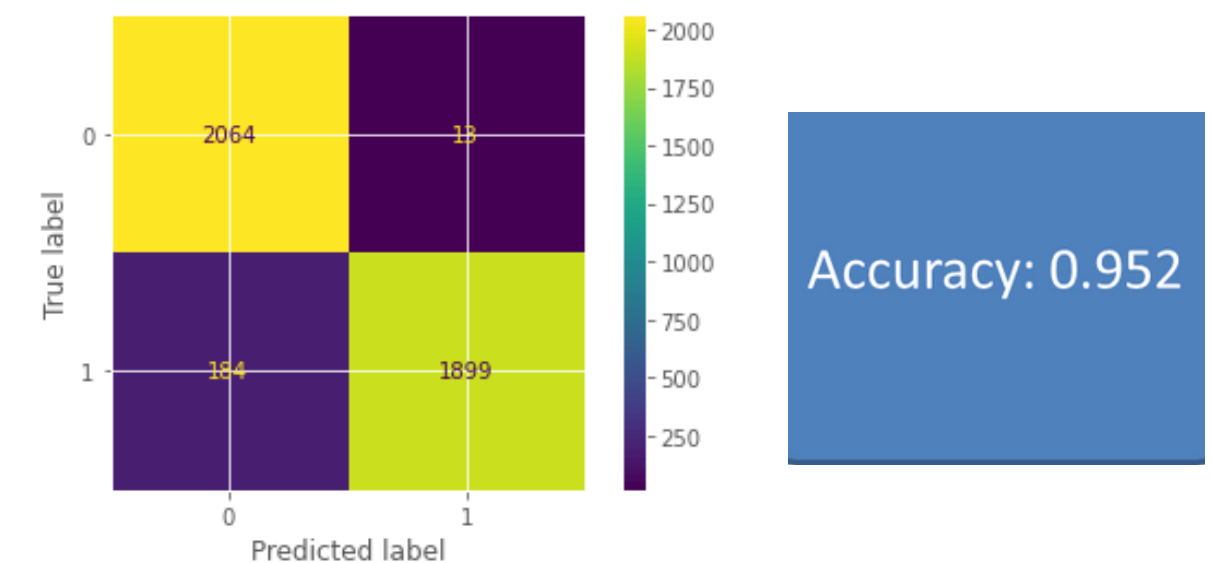
- Logistics regression is common and is a useful regression method for solving the binary classification problem.
- Logistic regression describes and estimates the relationship between one dependent binary variable and independent variables.



# Supervised models

## Multinomial Naive Bayes

- Multinomial Naive Bayes is one of the most popular supervised learning classifications that is used for the analysis of the categorical text data.
- Multinomial Naive Bayes algorithm is a probabilistic learning method that is mostly used in Natural Language Processing (NLP)

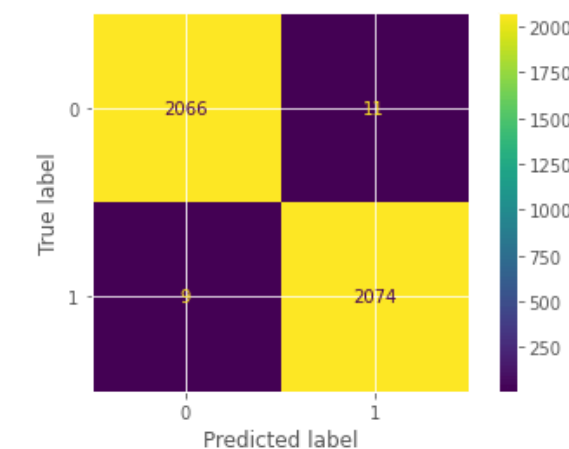




# Supervised models

## Decision Tree Classifier

- A decision tree is a supervised machine learning classification algorithm used to build models like the structure of a tree.
- In classification, the goal of the predictive model is to identify the class that generated a particular instance



Accuracy: 0.995



**Let's Test if it's is a Real or Fake**

THANK  
YOU



F A K E  
F A C T

