

NAME OF THE PROJECT

FAKE NEWS DETECTION

Submitted by:

Ankita Srivastava

ACKNOWLEDGMENT

This project is based on the NLP . I research on many site but I get some useful refrences which help me to make this project.

1-<https://www.javatpoint.com/nlp>

2-<https://www.educative.io/answers/preprocessing-steps-in-natural-language-processing-nlp>

3-<https://www.youtube.com/watch?v=5ctbvkAMQO4>

4-<https://www.youtube.com/watch?v=X2vAabgKiuM>

INTRODUCTION

Business Problem Framing:--

Fake news has become one of the biggest problems of our age. It has serious impact on our online as well as offline discourse. One can even go as far as saying that, to date, fake news poses a clear and present danger to western democracy and stability of the society.

Conceptual Background of the Domain Problem:---

There are two datasets one for fake news and one for true news. In true news, there is 21417 news, and in fake news, there is 23481 news. You have to insert one label column zero for fake news and one for true news. We are combined both datasets using pandas built-in function.

Analytical Problem Framing

Mathematical/ Analytical Modeling of the Problem

```
In [7]: df_fake_manual_testing = df_fake.tail(10)
        for i in range(23480,23470,-1):
            df_fake.drop([i], axis = 0, inplace = True)

        df_true_manual_testing = df_true.tail(10)
        for i in range(21416,21406,-1):
            df_true.drop([i], axis = 0, inplace = True)
```

Data sources and their format:--

Firstly I install many library because this programme need some installation like : worldcloud, nltk , xgboost, eli5.

The data set contains the fake dataset, which has approximately 23481 samples and the true dataset which contains nearly 21417 samples. All the data samples contain 5 fields which includes , "title", "text", "subject", "date", "id".

The class can be either 0 or 1, where 0 denotes a Fake while 1 denotes a True. There are various comments which have multiple labels. The first attribute is a unique ID associated with each comment.

Data Preprocessing Done:--

Data cleaning:

I have to check what is true and fake news.it divide in column.

```
In [8]: df_fake.shape , df_true.shape
```

```
Out[8]: ((23471, 5), (21407, 5))
```

```
In [9]: df_fake_manual_testing['class'] = 0  
df_true_manual_testing['class'] = 1
```

```
In [10]: df_fake_manual_testing.head(10)
```

```
Out[10]:
```

		title	text	subject	date	class
23471	Seven Iranians freed in the prisoner swap have...	21st Century Wire says This week, the historic...	Middle-east	January 20, 2016	0	
23472	#Hashtag Hell & The Fake Left	By Dady Chery and Gilbert MercierAll writers ...	Middle-east	January 19, 2016	0	
23473	Astroturfing: Journalist Reveals Brainwashing ...	Vic Bishop Waking TimesOur reality is carefull...	Middle-east	January 19, 2016	0	
23474	The New American Century: An Era of Fraud	Paul Craig RobertsIn the last years of the 20t...	Middle-east	January 19, 2016	0	
23475	Hillary Clinton: 'Israel First' (and no peace ...	Robert Fantina CounterpunchAlthough the United...	Middle-east	January 18, 2016	0	
23476	McPain: John McCain Furious That Iran Treated ...	21st Century Wire says As 21WIRE reported earl...	Middle-east	January 16, 2016	0	
23477	JUSTICE? Yahoo Settles E-mail Privacy Class-ac...	21st Century Wire says It s a familiar theme. ...	Middle-east	January 16, 2016	0	
23478	Sunnistan: US and Allied 'Safe Zone' Plan to T...	Patrick Henningsen 21st Century WireRemember ...	Middle-east	January 15, 2016	0	
23479	How to Blow \$700 Million: Al Jazeera America F...	21st Century Wire says Al Jazeera America will...	Middle-east	January 14, 2016	0	
23480	10 U.S. Navy Sailors Held by Iranian Military ...	21st Century Wire says As 21WIRE predicted in ...	Middle-east	January 12, 2016	0	