

Fake News Detection Using Machine Learning Algorithms In Python

Aim:-

The main goal of our project is to detect if the given news articles is **Fake** or **Real** using **Machine Learning Algorithms** in **Python** through **Jupyter notebook**.

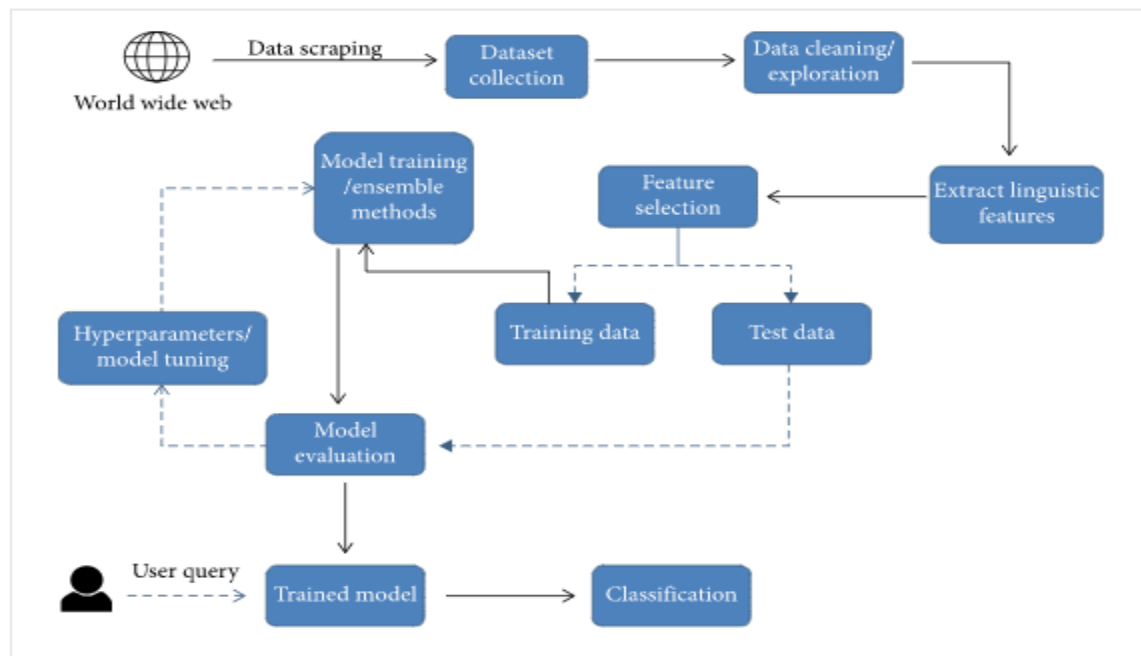
Abstract:-

Now a days, the use of internet is ubiquitous, everyone relies on various online resources for news. Along with the increase in the use of social media platforms like Facebook, Twitter, etc. news spread rapidly among millions of users within a very short span of time. The spread of fake news has far-reaching consequences like creation of biased opinions to swaying election outcomes for the benefit of certain candidates. So in this project we aim to detect the various news articles available online with the help of **Machine Learning Algorithms** and we aim to provide the users with the ability to classify the news as **Fake** or **Real** and also check the authenticity of the website publishing the news.

Methodology:-

This project is developed using Machine learning with Logistic Regression Prediction. This model takes the news articles and gives out the result that the given news article is a fake or real. This process is achieved by the manual testing, LR Prediction and Machine learning models for the result prediction with some small packages in Python.

Architecture:-



Software and Hardware Used:-

Software:-

Programming languages: Python

Machine Learning Packages: numpy

Pandas

train_test_split

classification_report

TfidfVectorizer

LogisticRegression

Hardware:-

Minimum of 4 GB RAM and a Jupyter

Technical Description:-

This project mainly consists of two phases:

1. In First Phase, we need to load the Datasets(**Fake.csv** and **True.csv**) from the disk and we are training the model using Logistic Regression Prediction on the Dataset.
2. In Second Phase once the model is trained, we are passing the input(news article) to the trained model and then it classifies whether the given news is **Fake** or **Real** and finally produces the appropriate result by doing manual testing the code by using the contents in manual_testing.csv.

Sourcecode:-

Fake_news_detection project using Logistic Regression

```
In [1]: import pandas as pd
import numpy as np
from sklearn.model_selection import train_test_split
from sklearn.metrics import classification_report
import re
import string
```

Datasets

```
In [2]: df_fake = pd.read_csv("Fake.csv")
df_true = pd.read_csv("True.csv")
```

```
In [3]: df_fake.head(10)
```

```
Out[3]:
```

	title	text	subject	date
0	Donald Trump Sends Out Embarrassing New Year'...	Donald Trump just couldn't wish all Americans ...	News	December 31, 2017
1	Drunk Bragging Trump Staffer Started Russian ...	House Intelligence Committee Chairman Devin Nu...	News	December 31, 2017
2	Sheriff David Clarke Becomes An Internet Joke...	On Friday, it was revealed that former Milwauk...	News	December 30, 2017
3	Trump Is So Obsessed He Even Has Obama's Name...	On Christmas day, Donald Trump announced that ...	News	December 29, 2017
4	Pope Francis Just Called Out Donald Trump Dur...	Pope Francis used his annual Christmas Day mes...	News	December 25, 2017
5	Racist Alabama Cops Brutalize Black Boy While...	The number of cases of cops brutalizing and ki...	News	December 25, 2017
6	Fresh Off The Golf Course, Trump Lashes Out A...	Donald Trump spent a good portion of his day a...	News	December 23, 2017
7	Trump Said Some INSANELY Racist Stuff Inside ...	In the wake of yet another court decision that...	News	December 23, 2017
8	Former CIA Director Slams Trump Over UN Bully...	Many people have raised the alarm regarding th...	News	December 22, 2017
9	WATCH: Brand-New Pro-Trump Ad Features So Muc...	Just when you might have thought we'd get a br...	News	December 21, 2017

```
In [4]: df_true.head(10)
```

```
Out[4]:
```

		title	text	subject	date
0		As U.S. budget fight looms, Republicans flip t...	WASHINGTON (Reuters) - The head of a conservat...	politicsNews	December 31, 2017
1		U.S. military to accept transgender recruits o...	WASHINGTON (Reuters) - Transgender people will...	politicsNews	December 29, 2017
2		Senior U.S. Republican senator: 'Let Mr. Muell...	WASHINGTON (Reuters) - The special counsel inv...	politicsNews	December 31, 2017
3		FBI Russia probe helped by Australian diplomat...	WASHINGTON (Reuters) - Trump campaign adviser ...	politicsNews	December 30, 2017
4		Trump wants Postal Service to charge 'much mor...	SEATTLE/WASHINGTON (Reuters) - President Donal...	politicsNews	December 29, 2017
5		White House, Congress prepare for talks on spe...	WEST PALM BEACH, Fla./WASHINGTON (Reuters) - T...	politicsNews	December 29, 2017
6		Trump says Russia probe will be fair, but time...	WEST PALM BEACH, Fla (Reuters) - President Don...	politicsNews	December 29, 2017
7		Factbox: Trump on Twitter (Dec 29) - Approval ...	The following statements were posted to the ve...	politicsNews	December 29, 2017
8		Trump on Twitter (Dec 28) - Global Warming	The following statements were posted to the ve...	politicsNews	December 29, 2017
9		Alabama official to certify Senator-elect Jone...	WASHINGTON (Reuters) - Alabama Secretary of St...	politicsNews	December 28, 2017

```
In [5]: df_fake["class"] = 0
df_true["class"] = 1
```

```
In [6]: df_fake.shape, df_true.shape
```

```
Out[6]: ((23481, 5), (21417, 5))
```

```
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)
```

```
In [8]: df_manual_testing = pd.concat([df_fake_manual_testing,df_true_manual_testing], axis = 0)
df_manual_testing.to_csv("manual_testing.csv")
```

```
In [9]: df_marge = pd.concat([df_fake, df_true], axis =0 )
df_marge.head(10)
```

```
Out[9]:
```

		title	text	subject	date	class
0		Donald Trump Sends Out Embarrassing New Year'...	Donald Trump just couldn t wish all Americans ...	News	December 31, 2017	0
1		Drunk Bragging Trump Staffer Started Russian ...	House Intelligence Committee Chairman Devin Nu...	News	December 31, 2017	0
2		Sheriff David Clarke Becomes An Internet Joke...	On Friday, it was revealed that former Milwauk...	News	December 30, 2017	0
3		Trump Is So Obsessed He Even Has Obama's Name...	On Christmas day, Donald Trump announced that ...	News	December 29, 2017	0
4		Pope Francis Just Called Out Donald Trump Dur...	Pope Francis used his annual Christmas Day mes...	News	December 25, 2017	0
5		Racist Alabama Cops Brutalize Black Boy While...	The number of cases of cops brutalizing and ki...	News	December 25, 2017	0
6		Fresh Off The Golf Course, Trump Lashes Out A...	Donald Trump spent a good portion of his day a...	News	December 23, 2017	0
7		Trump Said Some INSANELY Racist Stuff Inside ...	In the wake of yet another court decision that...	News	December 23, 2017	0
8		Former CIA Director Slams Trump Over UN Bully...	Many people have raised the alarm regarding th...	News	December 22, 2017	0
9		WATCH: Brand-New Pro-Trump Ad Features So Muc...	Just when you might have thought we d get a br...	News	December 21, 2017	0

```
In [10]: df = df_marge.drop(["title", "subject","date"], axis = 1)
df.head(10)
```

```
Out[10]:
```

		text	class
0		Donald Trump just couldn t wish all Americans ...	0
1		House Intelligence Committee Chairman Devin Nu...	0
2		On Friday, it was revealed that former Milwauk...	0
3		On Christmas day, Donald Trump announced that ...	0
4		Pope Francis used his annual Christmas Day mes...	0
5		The number of cases of cops brutalizing and ki...	0
6		Donald Trump spent a good portion of his day a...	0
7		In the wake of yet another court decision that...	0
8		Many people have raised the alarm regarding th...	0
9		Just when you might have thought we d get a br...	0

```
In [11]: df = df.sample(frac = 1)
```

```
In [12]: df.head()
```

```
Out[12]:
```

	text	class
15177	Godless heartless and without conscience. Sena...	0
18855	A producer at the Women in the World event had...	0
7159	(Reuters) - President-elect Donald Trump on Sa...	1
18949	Enough is enough. Americans need to start maki...	0
21107	This is a mind-blowing interview that confirms...	0

```
In [13]: df.isnull().sum()
```

```
Out[13]: text      0
class      0
dtype: int64
```

```
In [14]: def wordopt(text):
text = text.lower()
text = re.sub('[.*?\\]', '', text)
text = re.sub("\\W", "", text)
text = re.sub('https?://[S+|www\\.S+]', '', text)
text = re.sub('<.*?>+', '', text)
text = re.sub('[%s]' % re.escape(string.punctuation), '', text)
text = re.sub('\\n', '', text)
text = re.sub('\\w*\\d\\w*', '', text)
return text
```

```
In [15]: df["text"] = df["text"].apply(wordopt)
```

```
In [16]: df.head(10)
```

```
Out[16]:
```

	text	class
15177	godless heartless and without conscience sena...	0
18855	a producer at the women in the world event had...	0
7159	reuters president elect donald trump on sa...	1
18949	enough is enough americans need to start maki...	0
21107	this is a mind blowing interview that confirms...	0
7715	this october story has been refiled to say ...	1
3543	throughout the campaign many have tried defen...	0
11220	washington reuters the white house on frid...	1
6260	washington reuters u s president elect do...	1
10913	the students of the historically black bethune...	0

```
In [17]: x = df["text"]
y = df["class"]
```

```
In [18]: x_train, x_test, y_train, y_test = train_test_split(x, y, test_size=0.25)
```

```
In [19]: from sklearn.feature_extraction.text import TfidfVectorizer
```

```
In [20]: vectorization = TfidfVectorizer()
xv_train = vectorization.fit_transform(x_train)
xv_test = vectorization.transform(x_test)
```

Logistic Regression

```
In [21]: from sklearn.linear_model import LogisticRegression
```

```
In [22]: LR = LogisticRegression()
LR.fit(xv_train, y_train)
```

```
Out[22]: LogisticRegression()
```

```
In [23]: LR.score(xv_test, y_test)
```

```
Out[23]: 0.9861853832442068
```

```
In [24]: pred_LR=LR.predict(xv_test)
```

```
In [25]: print(classification_report(y_test, pred_LR))
```

	precision	recall	f1-score	support
0	0.99	0.98	0.99	5862
1	0.98	0.99	0.99	5358
accuracy			0.99	11220
macro avg	0.99	0.99	0.99	11220
weighted avg	0.99	0.99	0.99	11220

```
In [26]: def output_label(n):
        if n == 0:
            return "Fake News"
        elif n == 1:
            return "Not A Fake News"

        def manual_testing(news):
            testing_news = {"text": [news]}
            new_def_test = pd.DataFrame(testing_news)
            new_def_test["text"] = new_def_test["text"].apply(wordopt)
            new_x_test = new_def_test["text"]
            new_xv_test = vectorization.transform(new_x_test)
            pred_LR = LR.predict(new_xv_test)

            return print("\n\nLR Prediction: {} ".format(output_label(pred_LR[0])))
```

Result:-

The below picture is detecting the given news articles is whether **Fake or Not a fake news** using **Machine Learning Algorithms in Python with Logistic Regression(LR) Prediction**.

output

```
In [27]: news = str(input())
        manual_testing(news)

21st Century Wire says This week, the historic international Iranian Nuclear Deal was punctuated by a two-way prisoner swap between Washington and Tehran, but it didn't end quite the way everyone expected. On the Iranian side, one of the U.S. citizens who was detained in Iran, Nosratollah Khosravi-Roodsari, has stayed in Iran, but on the U.S. side all 7 of the Iranians held in U.S. prisons DID NOT show up to their flight to Geneva for the prisoner exchange with at least 3 electing to stay in the U.S. TEHRAN SIDE: In Iran, 5 U.S. prisoners were released, with 4 of them making their way to Germany via Switzerland. Will Robinson Daily Mail None of the Iranians freed in the prisoner swap have returned home and could still be in the United States, it has been reported. The seven former inmates, who were released as part of a deal with the Islamic republic, did not show up to get a flight to Geneva, Switzerland, where the exchange was set to take place on Sunday. Three of the Iranians have decided to stay in the United States, ABC reported, with some moving in with their families. However it is not known where the other four are. Three of the Americans who had been detained in Iran Washington Post journalist Jason Rezaian, former U.S. Marine Amir Hekmati and Christian pastor Saeed Abedini left Tehran at around 7am the same day, but weren't met by their counterparts in Switzerland. Continue this story at the Mail Online READ MORE IRAN NEWS AT: 21st Century Wire Iran Files

LR Prediction: Fake News
```

output

```
In [28]: news = str(input())
        manual_testing(news)

LONDON (Reuters) - LexisNexis, a provider of legal, regulatory and business information, said on Tuesday it had withdrawn two products from the Chinese market in March this year after it was asked to remove some content. The issue of academic freedom in China hit the headlines this week after the leading British academic publisher, Cambridge University Press, said it had complied with a request to block online access to some scholarly articles in China. It later reversed its position. Earlier this year LexisNexis Business Insight Solutions in China was asked to remove some content from its database, LexisNexis said in a statement. In March 2017, the company withdrew two products (Nexis and LexisNexis Academic) from the Chinese market. LexisNexis is owned by information group Relx.

LR Prediction: Not A Fake News
```

Developed by:-

P. Lakshmi Kumari : 208W5A0503

P. Vasanthi : 208W5A0504

Y. Lahari : 208W5A0506

V. Vema Sri : 208W5A0518