PROJECT REPORT

Fake News Detection

Course: Foundation of Data Science

Project Group Number: 13

Team member name:

Phaneendra AM.EN.U4CSE19332

A.Tharun kumar AM.EN.U4CSE19308

C Guru Prakash Reddy AM.EN.U4CSE19316

N Ram Chandra Sekhar AM.EN.U4CSE18238

S Sathwik Reddy AM.EN.U4CSE19366

Abstract:

we have a number of advantages of this digital world but it also has its disadvantages as well. There are different issues in this digital world. One of them is fake news. Someone can easily spread a fake news. Fake news is spread to harm the reputation of a person or an organization. It can be a propaganda against someone that can be a political party or an organization. There are different online platforms where the person can spread the fake news. This includes the Facebook, Twitter etc.

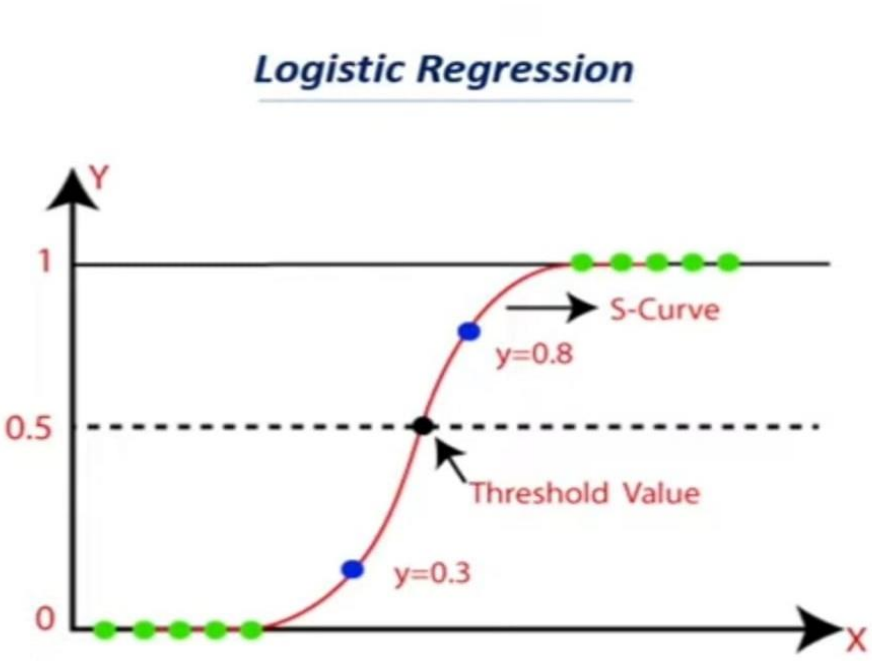
Introduction:

Internet is one of the important inventions and a large number of persons are its users. These persons use this for different purposes. There are different social media platforms that are accessible to these users. Any user can make a post or spread the news through these online platforms. These platforms do not verify the users or their posts. So some of the users try to spread fake news through these platforms. These fake news can be a propaganda against an

individual, society, organization or political party. A human being is unable to detect all these fake news. So there is a need for machine learning classifiers that can detect these fake news automatically. Use of machine learning classifiers for detecting the fake news is described in this systematic literature review.

Algorithm Used:

Logistic regression model: It is a supervised learning classification algorithm used to predict the probability of a target value.



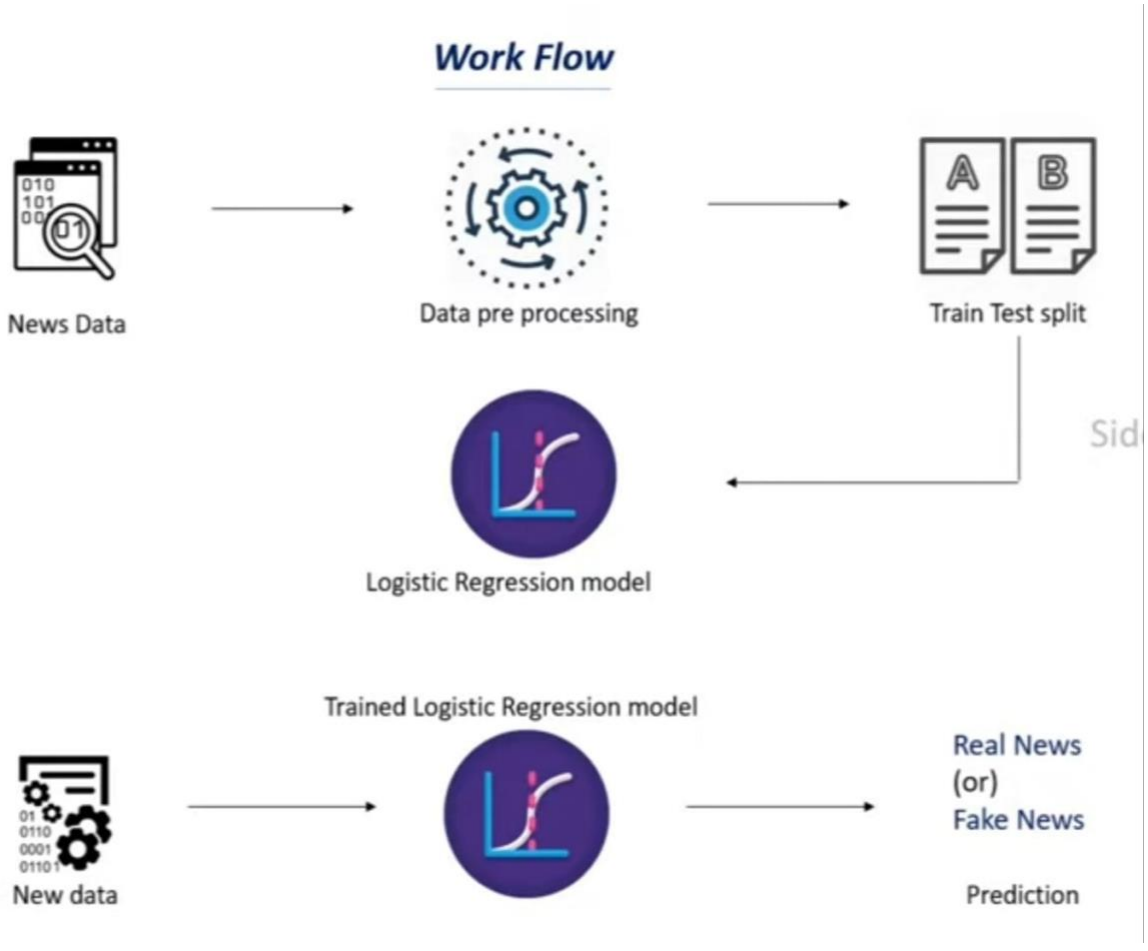
Methods:

We take the news given by the user as the input.

After, extraction of information takes place from dataset attributes.

Now, by using algorithm we will be identifying the news is real or fake news.

Finally, the news real or fake is displayed.



Dataset:

We used train dataset from kaggle. We refer notes from kaggle https://www.kaggle.com/clmentbisaillon/fake-and-real-news-dataset.

The dataset includes the following attributes: id no (unique id for a news article); news_title (title of a news article); author of news (author of the news article); news_text (text of the news article); label (whether the news is real or fake: 1 represent fake news, 0 represent real news).

Data pre-processing:

loading the dataset from pandas

```
#Data Pre-processing
news = pd.read_csv('news_dataset.csv')          # loading the dataset to a pandas DataFrame
```

printing first five rows of the dataset

```
news.head()          # printing the first 5 rows of the dataframe
```

	id no	news_title	author of news	news_text	label
0	0	House Dem Aide: We Didn't Even See Comey's Let...	Darrell Lucus	House Dem Aide: We Didn't Even See Comey's Let...	1
1	1	FLYNN: Hillary Clinton, Big Woman on Campus - ...	Daniel J. Flynn	Ever get the feeling your life circles the rou...	0
2	2	Why the Truth Might Get You Fired	Consortiumnews.com	Why the Truth Might Get You Fired October 29, ...	1
3	3	15 Civilians Killed In Single US Airstrike Hav...	Jessica Purkiss	Videos 15 Civilians Killed In Single US Airstr...	1
4	4	Iranian woman jailed for fictional unpublished...	Howard Portnoy	Print \nAn Iranian woman has been sentenced to...	1

counting the number of missing values in the dataset

```
news.isnull().sum()          #counting the number of missing values in the dataset
```

```
id no               0
news_title        558
author of news   1957
news_text          39
label               0
dtype: int64
```

Data Split: Split the dataset into x_train with 80% content and

x_test with 20% content.

Splitting the dataset to training & test data

```
X_train, X_test, Y_train, Y_test = train_test_split(X, Y, test_size = 0.2, stratify=Y, random_state=2)
```

Accuracy score:

Accuracy Score

```
X_train_prediction = model.predict(X_train)              # accuracy score on the training data
training_data_accuracy = accuracy_score(X_train_prediction, Y_train)
```

```
print('Accuracy score of the training data : ', training_data_accuracy)
```

```
Accuracy score of the training data :  0.9865985576923076
```

```
# accuracy score on the test data
X_test_prediction = model.predict(X_test)
test_data_accuracy = accuracy_score(X_test_prediction, Y_test)
```

```
print('Accuracy score of the test data : ', test_data_accuracy)
```

```
Accuracy score of the test data :  0.9790865384615385
```

Result:

```
X_new = X_test[3]

prediction = model.predict(X_new)
print(prediction)

if (prediction[0]==0):
  print('The news is Real')
else:
  print('The news is Fake')
```

```
[0]
The news is Real
```

Conclusion:

Due to increasing use of internet, it is now easy to spread fake news. A huge number of persons are regularly connected with internet and social media platforms. There is no any restriction while posting any news on these platforms. There is a need for a way to detect these fake news by using classifiers.