

APPENDIX 1

FAKE NEWS DETECTION

END TERM REPORT

By

Name: Ushkamalla kotes h

Section: KM029

Roll Number: RKM029B60



L LOVELY
P ROFESSIONAL
U NIVERSITY

Transforming Education Transforming India

Department of Intelligent Systems

School of Computer Science Engineering

Lovely Professional University, Jalandhar

November, 2022

APPENDIX 2

Student Declaration

This is to declare that this report has been written by me/us. No part of the report is copied from other sources. All information included from other sources have been duly acknowledged. I/We aver that if any part of the report is found to be copied, I/we shall take full responsibility for it.

Name: Ushkamalla kotesb

Roll number:60

APPENDIX 4

BONAFIDE CERTIFICATE

Certified that this project report FAKE NEWS DETECTION is the Bonafede work of USHKAMALLA KOTESH who carried out the project work under my supervision.

DR DHANPRATAP SINGH

Abstract:

Now-a-days we generally listen that many fake and false news about many celebrities, politicians etc .Widespread belief in misinformation circulating online is a critical challenge for modern societies. According to the study falsehood diffuses significantly farther, faster, deeper, and more broadly than the truth, in all categories of information, and in many cases by an order of magnitude. This is done mainly by click baits which lure users and entice curiosity with flashy headlines or designs and trick them into clicking the links to increase the ad revenues. Hence it is crucial to determine the integrity of the information available on the internet in order to keep the prevalence of fake news in check.

The development of the internet and technology have led the world to online, every single step and every single move has become online, it seems everything can be done just with a click on the internet. It is not just limited to buying groceries or booking tickets, it has been exaggerated to online video meetings, online learning and more and more. Everyone now relies on several online news sources because the internet is so pervasive in our modern world. In addition to the rising popularity of social media sites like Facebook, Twitter, etc., the news quickly reached millions of people in a short period. The propagation of misleading information has wide-ranging impacts, such as the development of ideologies that are skewed in favour of particular politicians. Spammers also monetize ads by clicking on barriers and using alluring news headlines. Online forums are where the majority of smartphone users choose to read the news. News websites provide breaking news and act as a source of authority. The issue is how to deliver news and articles on social media platforms like WhatsApp groups, Facebook pages, Twitter, and other little blogs and social networking sites. To spread these stories and produce news, the public runs the risk of harm. There is an urgent need to put an end to rumours, especially in growing nations like India, and to concentrate on true, established issues. Fake news has spread around the world since the emergence of the media. People are now distrustful of bogus news as a result. In todays digital environment, when there are countless forums where false news or incorrect information can propagate, the pervasive issue of fake news is one of the most challenging to address. The issue of artificial bots that can be used to fabricate and disseminate lies being brought about by the development of Artificial Intelligence is significant news. The majority of people believe everything they read online, and those who lack literacy or are unfamiliar with digital technologies can easily be duped, which makes the situation tense. Fraudulent spam or malware emails and texts may cause the same issue. As a result, it is necessary to acknowledge this issue in order to tackle the challenge of reducing crime, political turmoil, misery, and attempts to spread false information.

This paper aims to develop the fake news detector which is able to detect any kinds of fake news. In order to detect the fake news, From the experimental results, this device is able to detect the fake news.

So we will try to build a simple Machine Learning Model using Logistic Regression to detect whether a news article is fake or not.

INTRODUCTION:

Information is readily available thanks to the rising use of social media and other mobile technologies. For the dissemination of news and information, social media platforms and mobile applications have displaced traditional print media. People naturally exhibit a great desire to use digital media for their daily information demands given the comfort and speed it offers. In addition to giving customers rapid access to a range of data, it also gives for-profit organisations a solid platform for reaching a larger audience. It appears tedious for the forum to distinguish between actual news and bogus news in terms of information. False information is frequently spread with the aim of deceiving people or fostering prejudice to benefit from it politically or financially. As a result, it might include interesting news items or other content to draw in more users. The veracity of different news reports that favoured candidates and their political agendas during the most recent US elections has been hotly contested. The investigation of fake news is gaining traction in the face of this growing concern in an effort to stop its damaging impacts on people and communities. Machine learning algorithms including Vector Support Machines, Random Forests, Decision Trees, , and others are frequently utilised by fake news detection systems. In this project, we must put into practice a model that uses a logistic regressor classifier to categorise news as authentic or phoney.

The primary objective of the proposed work is to detect the fake news to ensure creditability, benefits of the real news, to deserve the truth by using Machine learning. Recent elections in the United States and other countries expose the creation of fake stories, which are oftentimes spread in an attempt to sway students' political opinions or worldviews. False information is spread across all forums and can originate from a wide range of sources. The fact that the information appears to have been created by respected news organisations is one of the traits of false news. It becomes increasingly more difficult for news reporters to determine what is accurate as a result of additional false information kinds including deepfakes, biased reporting, and sources that are only partially mentioned. The majority of the recent rumours about social media include social media, even though fake news is not a new issue and is present in all media sources, including books, TV, radio, and the Internet. Despite the efforts of numerous companies to locate and eliminate them, false news frequently circulates on social media sites.

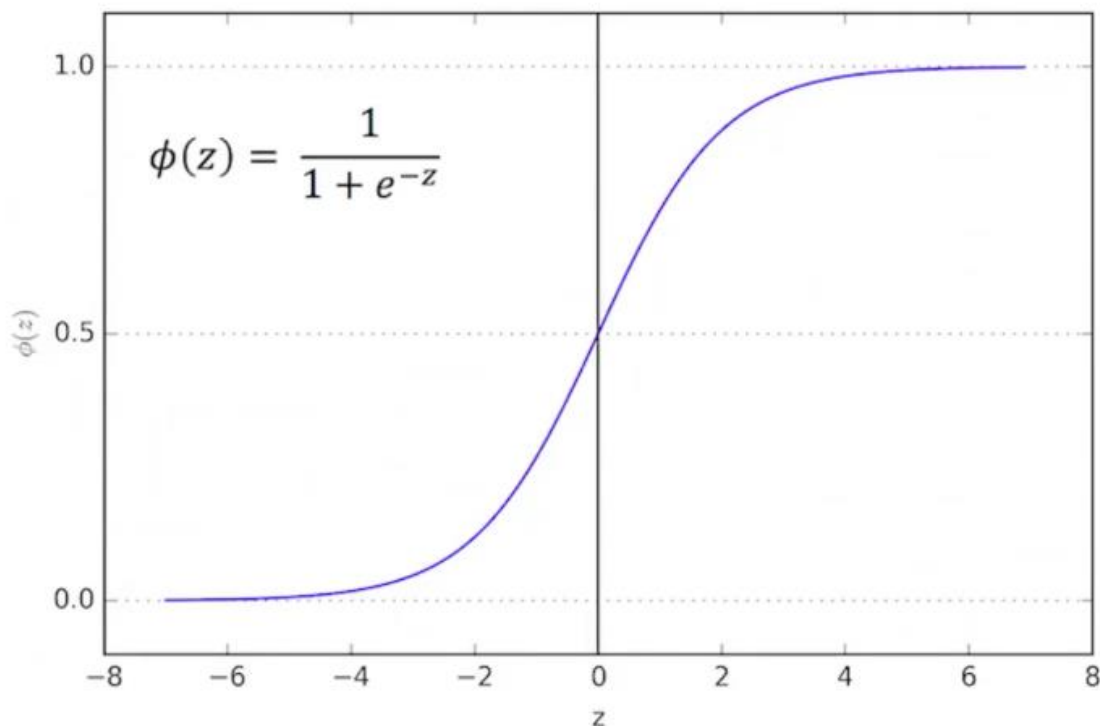
Some consumers of news continue to worry about the calibre of the content they see on these websites. For instance, older generations are less trusting in news on social media than younger generations, according to a survey of news consumers conducted annually. However, for other people, their response to the news does not appear to be impacted by this lack of confidence. A region of purchasers declares that one in all their favoured sports on social media is analysing or looking the information. Compared to sure older information purchasers, Gen Z and millennials are much more likely to call unique famous social media systems as one in all their foremost reasserts of information and information. They also don't express as much mistrust of social media. In this proposed effort, we will introduce a new framework for the detection of false information, called fake news detection, to address the problems. The suggested model in this study attempts to learn to forecast in order to simultaneously infer the trustworthiness labels of news pieces, creators, and subjects. The fake news detection challenge is constructed on the premise of the loyalty points problem.

Description of Project:

I downloaded the dataset from the Kaggle and imported. after importing libraries and the dataset I performed data pre-processing step, it is important to preprocess the data before we train our ML model since there might be some anomalies and missing datapoints which might make our predictions a bit skewed from the actual values.

Logistic Regression:

Logistic regression is a statistical analysis method to predict a binary outcome, such as yes or no(binary classification), based on prior observations of a data set. It is a Supervised statistical technique to find the probability of dependent variable. The graph shown below is a Sigmoid Function, which we also call as a Logit. This function converts the probabilities into binary values which could be further used for predictions.



According to this graph, if we obtain the probability value to be less than 0.5, then it is considered to be of the Class 0 and if the value is more than 0.5, then it would be a part of Class 1.

Libraries used in the project:

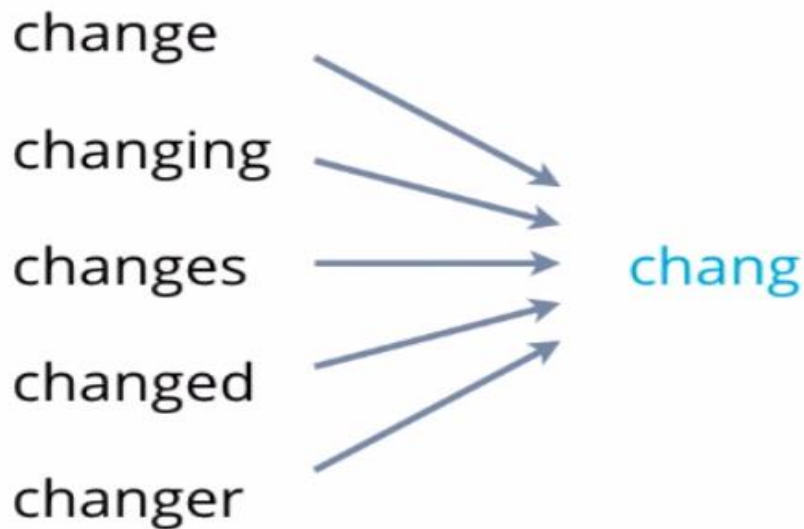
The importance of each library/module/function that we imported:

1. **NumPy:** It is a general-purpose array and matrices processing package.
2. **Pandas:** It allows us to perform various operations on datasets.
3. **re:** It is a built-in RegEx package, which can be used to work with Regular Expressions.
4. **NLTK:** It is a suite of libraries and programs for symbolic and statistical natural language processing (NLP).
5. **Nltk.corpus:** This package defines a collection of corpus reader classes, which can be used to access the contents of diverse set of corpora.
6. **Stopwords:** The words which are generally filtered out before processing a natural language are called stop words. These are actually the most common words in any language (like articles, prepositions, pronouns, conjunctions, etc) and does not add much information to the text. (Example-and, of, are etc.)
7. **PosrterStemmer:** A package to help us with stemming of words.
8. **Sci-kit Learn (sklearn) :** It provides a selection of efficient tools for machine learning and statistical modelling including classification, regression, clustering and dimensionality reduction via a consistence interface in Python.
9. **Feature_extraction.text:** It is used to extract features in a format supported by machine learning algorithms from datasets consisting of text.
10. **TfidfVecorizer:** It transforms text to feature vectors that can be used as input to estimator.
11. **train_test+split:** It is a function in Sklearn model selection for splitting data arrays into two subsets - for training data and for testing data.
12. **metrics** and **accuracy_score :** To import Accuracy classification score from the metrics module.

Stemming:

Stemming is the process of reducing a word to its word stem that affixes to suffixes and prefixes or to the roots of words. Stemming is the process of producing morphological variants of a root/base word. Stemming programs are commonly referred to as stemming algorithms or stemmers. A stemming algorithm reduces the words “chocolates”, “chocolatey”, “choco” to the root word, “chocolate” and “retrieval”, “retrieved”, “retrieves” reduce to the stem “retrieve”. Stemming is an important part of the pipelining process in Natural language processing.

Below figure shows the example for stemming.



```
In [44]: 1 data = pd.read_csv('fakenews.csv')
         2 data.head()
```

```
Out[44]:
```

	id	title	author	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 Aistr...	1
4	4	Iranian woman jailed for fictional unpublished...	Howard Portnoy	Print \nAn Iranian woman has been sentenced to...	1

```
In [45]: 1 data.shape
```

```
Out[45]: (20800, 5)
```

From the above screenshot we observe that importing the dataset and the shape of the dataset is 20800,5 it means


```
In [56]: 1 model = LogisticRegression()
         2 model.fit(X_train, y_train)
```

```
Out[56]: LogisticRegression()
```

```
In [29]: 1 X_train_prediction = model.predict(X_train)
         2 training_accuracy = accuracy_score(X_train_prediction, y_train)
         3 print(training_accuracy)
```

```
0.9865985576923076
```

```
In [30]: 1 X_test_prediction = model.predict(X_test)
         2 testing_accuracy = accuracy_score(X_test_prediction, y_test)
         3 print(testing_accuracy)
```

```
0.9790865384615385
```

We trained the model using logistic regressor and we got accuracy 98 percent in training set data and we got 97 % accuracy from testing set data.

	id	title	author	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 Aistr...	1
4	4	Iranian woman jailed for fictional unpublished...	Howard Portnoy	Print \nAn Iranian woman has been sentenced to...	1

From the dataset we can observe that the column label indicates it is a fake news or original news it is used for evaluation for the model in the label column value 1 represent it is a fake and the value 0 represent the news is not fake.

```
In [35]: 1 prediction = model.predict(X_sample)
2 if prediction == 0:
3     print('The NEWS is Real!')
4 else:
5     print('The NEWS is Fake!')
```

The NEWS is Fake!

```
In [36]: 1 X2_sample=X_test[1]
```

```
In [37]: 1 prediction = model.predict(X2_sample)
2 if prediction == 0:
3     print('The NEWS is Real!')
4 else:
5     print('The NEWS is Fake!')
```

The NEWS is Real!

```
In [41]: 1 X3_sample=X_test[2]
```

```
In [42]: 1 prediction = model.predict(X3_sample)
2 if prediction == 0:
3     print('The NEWS is Real!')
4 else:
5     print('The NEWS is Fake!')
```

The NEWS is Fake!

Technologies and frameworks:

1. Jupiter notebook
2. Natural Language Processing
3. Deep Learning
4. Stemming

CONCLUSION:

Newspapers that were earlier preferred as hard- copies are now being substituted by applications like Facebook, Twitter, and news articles to be read online. Whatsapps forwards are also a major source. The growing problem of fake news only makes things more complicated and tries to change or hamper the opinion and attitude of people towards use of digital technology. When a person is deceived by the real news two possible things happen- People start believing that their perceptions about a particular topic are true as assumed. Thus, in order to curb the phenomenon, we have developed our Fake news Detection system that takes input from the user and classify it to be true or fake. To implement this, various NLP and Machine Learning Techniques have to be used. The model is trained using an appropriate dataset and performance evaluation is also done using various performance measures. The best model, i.e., the model with highest accuracy is used to classify the news headlines or articles. As evident above for staticsearch, our best model came out to be Logistic Regression with an accuracy of 65%. Hence, we then used grid search parameter

optimization to increase the performance of logistic regression which then gave us the accuracy of 75%. Hence, we can say that if a user feed a particular news article or its headline in our model, there are 75% chances that it will be classified to its true nature.

The user can check the news article or keywords online; he can also check the authenticity of the website. The accuracy for dynamic system is 93% and it increases with every iteration.

References:

- <https://www.kaggle.com>
- <https://medium.com>

Git-Hub: [kotes1234/Fake-News-Detection \(github.com\)](https://github.com/kotes1234/Fake-News-Detection)