A Mini Project Report

On

Fake News Detection

By

Vaishnavi Sambhaji Sakhare (72007323F)
Abhay Tejbahadur Vishwakarma (72014424J)

Under the Guidance Of

Prof. Chetana Baviskar Ma'am

Department of Computer Engineering

Alard College of Engineering and Management, Marunji, Pune

SAVITRIBAI PHULE PUNE UNIVERSITY

2021-2022

Department of Computer Engineering
Alard College of Engineering and
Management, Pune.

Date: 16/12/2021

CERTIFICATE

This is to certify that,

Vaishnavi Sambhaji Sakhare(72007323F)
Abhay Tejbahadur Vishwakarma(72014424J)

of class B.E Computer; have successfully completed their mini project work on “Fake news detection” at Alard College of Engineering and Management, Pune, in the partial fulfillment of the Graduate Degree course in B.E at the department of **Computer Engineering** in the academic Year 2021-2022 Semester-I as prescribed by the Savitribai Phule Pune University.

Project Guide

Head of Department

ACKNOWLEDGEMENT

With deep sense of gratitude, we would like to thanks all the people who have lit our path with their kind guidance. We are very grateful to these intellectuals who did their best to help during our project work.

It is our proud privilege to express deep sense of gratitude to **Dr. K. D. Sapate,** Principal of Alard College of Engineering and Management, Pune, for his comments and kind permission to complete this project. We remain indebted to **Prof. Priyadarshani Doke**, H.O.D. of Computer Engineering Department for her timely suggestion and valuable guidance.

The special gratitude goes to (**Prof. Chetana Baviskar**) excellent and precious guidance in completion of this work. We thanks to all the colleagues for their appreciable help for our working project. With various industry owners or lab technicians to help, it has been our endeavor to throughout our work to cover the entire project work.

We also thankful to our parents who providing their wishful support for our project completion successfully. And lastly, we thanks to our all friends and the people who are directly or indirectly related to our project work.

Vaishnavi Sambhaji Sakhare (72007323F)
Abhay Tejbahadur Vishwakarma(72014424J)

Contents

ABSTRACT

Abstract: In our modern era where the 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 the creation of biased opinions to saying election outcomes for the benefit of certain candidates. Moreover, spammers use appealing news headlines to generate revenue using advertisements via clickbait's. In this paper, we aim to perform binary classification of various news articles available online with the help of concepts pertaining to Artificial Intelligence, Natural Language Processing and Machine Learning.

We aim to provide the user with the ability to classify the news as fake or real and also check the authenticity of the website publishing the news.

INTRODUCTION

As an increasing amount of our lives is spent interacting online through social media platforms, more and more people tend to hunt out and consume news from social media instead of traditional news organizations. The explanations for this alteration in consumption behaviors are inherent within the nature of those social media platforms: (i) it's often more timely and fewer expensive to consume news on social media compared with traditional journalism, like newspapers or television; and (ii) it's easier to further share, discuss , and discuss the news with friends or other readers on social media. For instance, 62 percent of U.S. adults get news on social media in 2016, while in 2012; only 49 percent reported seeing news on social media. It had been also found that social media now outperforms television because the major news source. Despite the benefits provided by social media, the standard of stories on social media is less than traditional news organizations. it had been estimated that over 1 million tweets are associated with fake news —Pizzagate" by the top of the presidential election. The extensive spread of faux news can have a significant negative impact on individuals and society. Second, fake news intentionally persuades consumers to simply accept biased or false beliefs. It's crucial that we build up methods to automatically detect fake news broadcast on social media. Internet and social media have made the access to the news information much easier and comfortable.

METHODOLOGY

The system which is developed in three parts. The first part is static which works on machine learning classifier. We studied and trained the model with 4 different classifiers and chose the best classifier for final execution. The second part is dynamic which takes the keyword/text from user and searches online for the truth probability of the news. The third part provides the authenticity of the URL input by user. In this paper, we have used Python and its Sci-kit libraries. Python has a huge set of libraries and extensions, which can be easily used in Machine Learning. Sci-Kit Learn library is the best source for machine learning algorithms where nearly all types of machine learning algorithms are readily available for Python, thus easy and quick evaluation of ML algorithms is possible. We have used Django for the web-based deployment of the model, provides client-side implementation using HTML, CSS and JavaScript. We have also used Beautiful Soup (bs4), requests for online scrapping.

IMPLEMENTATION

4.1 DATA COLLECTION AND ANALYSIS: -

We can get online news from different sources like social media websites, search engine, homepage of news agency websites or the fact-checking websites These datasets have been widely used in different research papers for determining the veracity of news. In the following sections, I have discussed in brief about the sources of the dataset used in this work. Online news can be collected from different sources, such as news agency homepages, search engines, and social media websites. However, manually determining the veracity of news is a challenging task, usually requiring annotators with domain expertise who performs careful analysis of claims and additional evidence, context, and reports from authoritative sources. Generally, news data with annotations can be gathered in the following ways: Expert journalists, Fact-checking websites, Industry detectors, and Crowd sourced workers. However, there are no agreed upon benchmark datasets for the fake news detection problem. The dataset that we used is explained below:

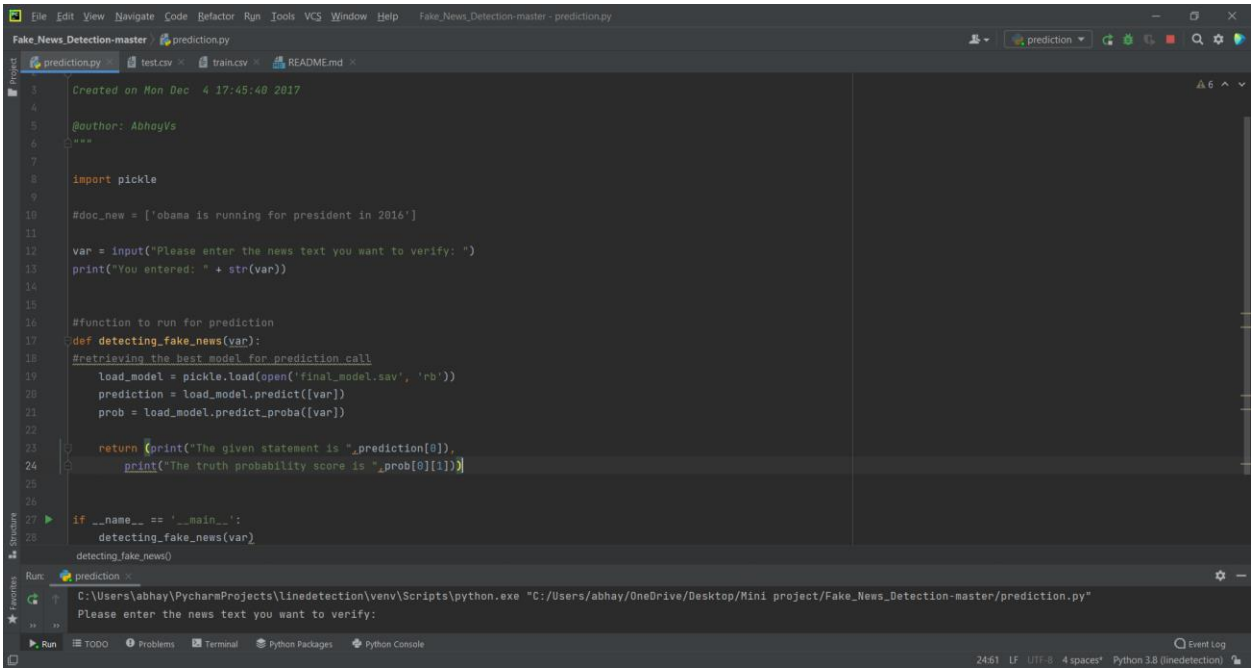
LIAR: The labels for news truthfulness are fine-grained multiple classes: pants-fire, false, barely-true, half-true, mostly true, and true. The data source used for this project is LIAR dataset which contains 3 files with .csv format for test, train and validation. Below is some description about the data files used for this project.

1. LIAR: Below are the columns used to create 3 datasets that have been in used in this project-
● Column1: Statement (News headline or text).

● Column2: Label (Label class contains: True, False) The dataset used for this project were in csv format named train.csv, test.csv and valid.csv. 2.REAL_OR_FAKE.CSV we used this dataset for passive aggressive classifier. It contains 3 columns viz 1- Text/keyword, 2-Statement, 3-Label (Fake/True)

SOURCE CODE AND OUTPUT OF FAKE NEWS DETECTION SYSTEM

❖ SOURCE CODE:



❖ OUTPUT:-

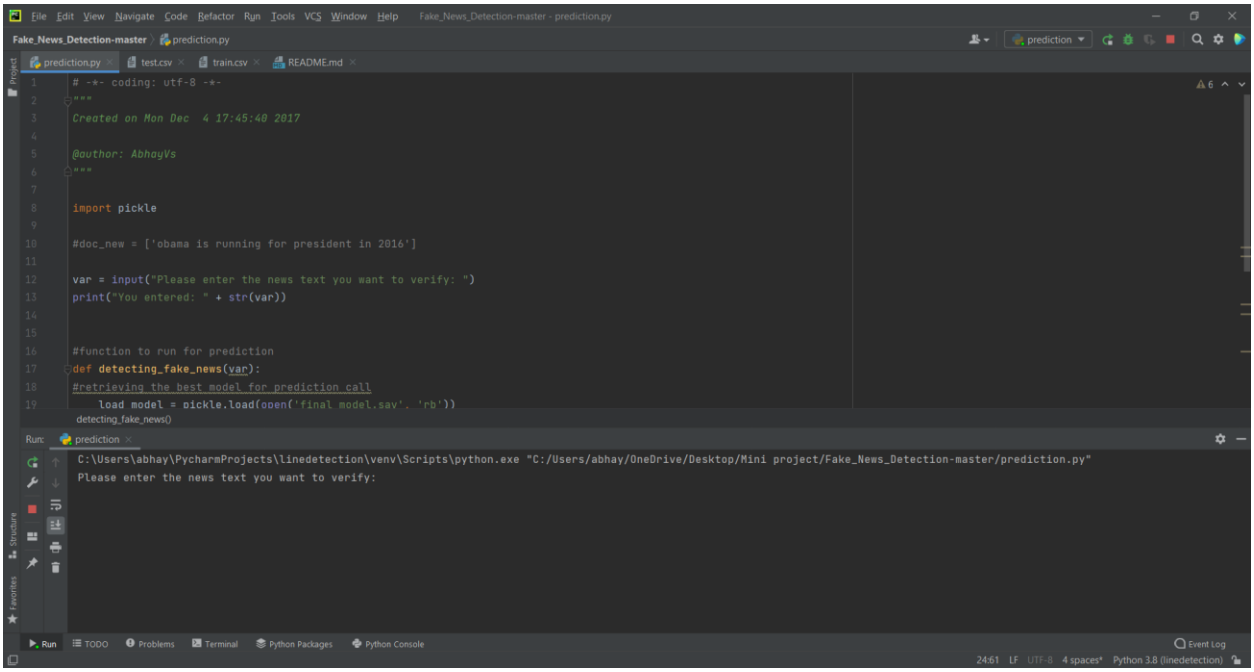


Fig: a

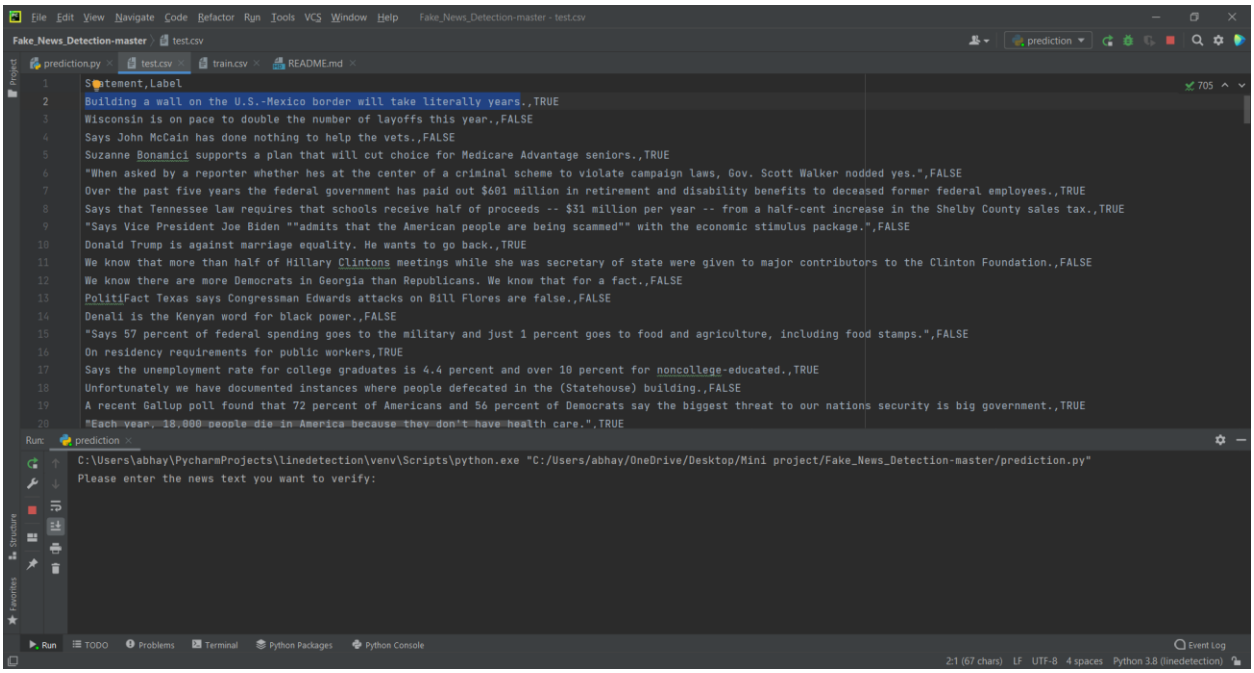

Statement,Label
Building a wall on the U.S.-Mexico border will take literally years.,TRUE
Wisconsin is on pace to double the number of layoffs this year.,FALSE
Says John McCain has done nothing to help the vets.,FALSE
Suzanne Bonamici supports a plan that will cut choice for Medicare Advantage seniors.,TRUE
"When asked by a reporter whether hes at the center of a criminal scheme to violate campaign laws, Gov. Scott Walker nodded yes.",FALSE
Over the past five years the federal government has paid out $601 million in retirement and disability benefits to deceased former federal employees.,TRUE
Says that Tennessee law requires that schools receive half of proceeds -- $31 million per year -- from a half-cent increase in the Shelby County sales tax.,TRUE
"Says Vice President Joe Biden ""admits that the American people are being scammed"" with the economic stimulus package.",FALSE
Donald Trump is against marriage equality. He wants to go back.,TRUE
We know that more than half of Hillary Clintons meetings while she was secretary of state were given to major contributors to the Clinton Foundation.,FALSE
We know there are more Democrats in Georgia than Republicans. We know that for a fact.,FALSE
PolitiFact Texas says Congressman Edwards attacks on Bill Flores are false.,FALSE
Denali is the Kenyan word for black power.,FALSE
"Says 57 percent of federal spending goes to the military and just 1 percent goes to food and agriculture, including food stamps.",FALSE
On residency requirements for public workers,TRUE
Says the unemployment rate for college graduates is 4.4 percent and over 10 percent for noncollege-educated.,TRUE
Unfortunately we have documented instances where people defecated in the (Statehouse) building.,FALSE
A recent Gallup poll found that 72 percent of Americans and 56 percent of Democrats say the biggest threat to our nations security is big government.,TRUE
"Each year, 18,000 people die in America because they don't have health care.",TRUE
C:\Users\abhay\PycharmProjects\linedetection\venv\Scripts\python.exe "C:/Users/abhay/OneDrive/Desktop/Mini project/Fake_News_Detection-master/prediction.py"
Please enter the news text you want to verify:

Fig:b

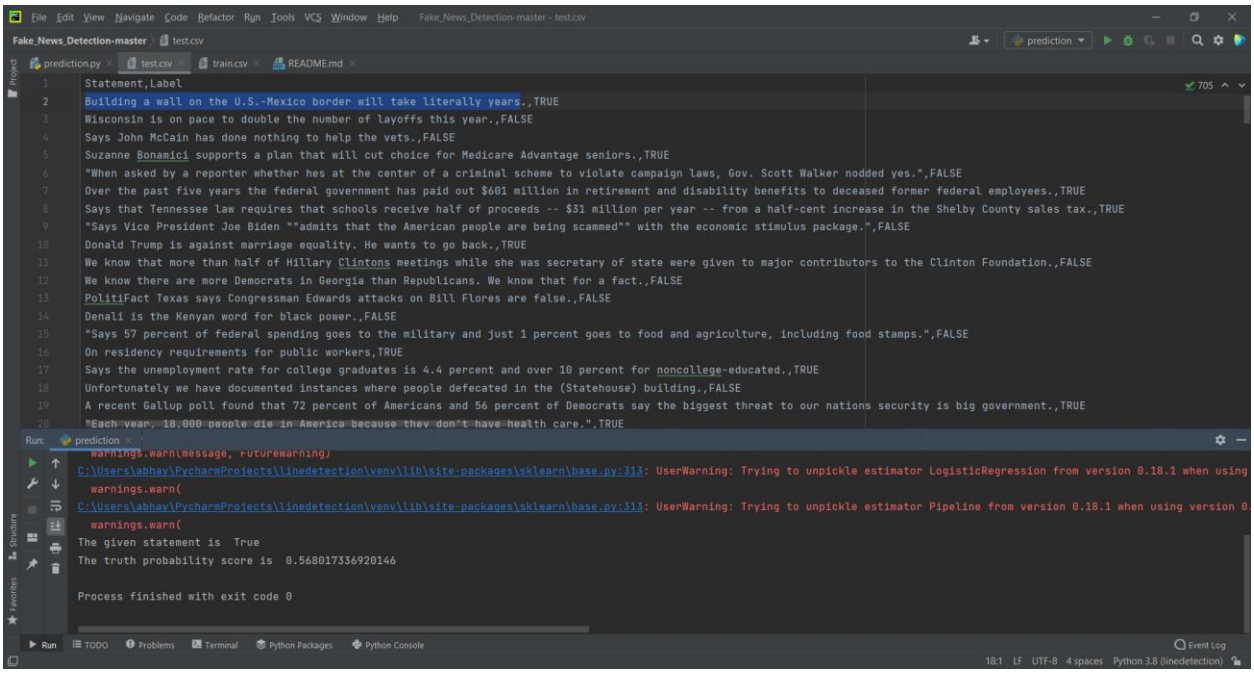

Statement,Label
Building a wall on the U.S.-Mexico border will take literally years.,TRUE
Wisconsin is on pace to double the number of layoffs this year.,FALSE
Says John McCain has done nothing to help the vets.,FALSE
Suzanne Bonamici supports a plan that will cut choice for Medicare Advantage seniors.,TRUE
"When asked by a reporter whether hes at the center of a criminal scheme to violate campaign laws, Gov. Scott Walker nodded yes.",FALSE
Over the past five years the federal government has paid out $601 million in retirement and disability benefits to deceased former federal employees.,TRUE
Says that Tennessee law requires that schools receive half of proceeds -- $31 million per year -- from a half-cent increase in the Shelby County sales tax.,TRUE
"Says Vice President Joe Biden ""admits that the American people are being scammed"" with the economic stimulus package.",FALSE
Donald Trump is against marriage equality. He wants to go back.,TRUE
We know that more than half of Hillary Clintons meetings while she was secretary of state were given to major contributors to the Clinton Foundation.,FALSE
We know there are more Democrats in Georgia than Republicans. We know that for a fact.,FALSE
PolitiFact Texas says Congressman Edwards attacks on Bill Flores are false.,FALSE
Denali is the Kenyan word for black power.,FALSE
"Says 57 percent of federal spending goes to the military and just 1 percent goes to food and agriculture, including food stamps.",FALSE
On residency requirements for public workers,TRUE
Says the unemployment rate for college graduates is 4.4 percent and over 10 percent for noncollege-educated.,TRUE
Unfortunately we have documented instances where people defecated in the (Statehouse) building.,FALSE
A recent Gallup poll found that 72 percent of Americans and 56 percent of Democrats say the biggest threat to our nations security is big government.,TRUE
"Each year, 18,000 people die in America because they don't have health care.",TRUE
warnings.warn(message, FutureWarning)
C:\Users\abhay\PycharmProjects\linedetection\venv\lib\site-packages\sklearn\base.py:313: UserWarning: Trying to unpickle estimator LogisticRegression from version 0.18.1 when using
warnings.warn(
C:\Users\abhay\PycharmProjects\linedetection\venv\lib\site-packages\sklearn\base.py:313: UserWarning: Trying to unpickle estimator Pipeline from version 0.18.1 when using version 0.
warnings.warn(
The given statement is True
The truth probability score is 0.568017336920146
Process finished with exit code 0

Fig: c

INSTALLATION REQUIREMENTS

- **Hardware and Software Requirements: -**

  - **Hardware Requirements: -**

    - Minimum 1 GB RAM.
    - Minimum 120 GB HDD
    - Processor Pentinium4 & Above

  - **Software Requirement: -**

    - Windows 7/10/11 and above
    - PyCharm Editor

Conclusion

In the 21st century, the majority of the tasks are done online. Newspapers that were earlier preferred as hardcopies are now being substituted by applications like Facebook, Twitter, and news articles to be read online. WhatsApp's 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

REFERENCES

[1] Kai Shu, Amy Sliva, Suhang Wang, Jiliang Tang, and Huan Liu, —Fake News Detection on Social Media: A Data Mining Perspective‖ arXiv:1708.01967v3 [cs.SI], 3 Sep 2017

[2] Kai Shu, Amy Sliva, Suhang Wang, Jiliang Tang, and Huan Liu, —Fake News Detection on Social Media: A Data Mining Perspective‖ arXiv:1708.01967v3 [cs.SI], 3 Sep 2017

[3] M. Granik and V. Mesyura, "Fake news detection using naive Bayes classifier," 2017 IEEE First Ukraine Conference on Electrical and Computer Engineering (UKRCON), Kiev, 2017, pp. 900-903.

[4] Fake news websites. (n.d.) Wikipedia. [Online]. Available: https://en.wikipedia.org/wiki/Fake_news_website. Accessed Feb. 6, 2017

[5] Cade Metz. (2016, Dec. 16). The bittersweet sweepstakes to build an AI that destroys fake news.

[6] Conroy, N., Rubin, V. and Chen, Y. (2015).—Automatic deception detection: Methods for finding fake news‖ at Proceedings of the Association for Information Science and Technology, 52(1), pp.1-4.

[7] Markines, B., Cattuto, C., & Menczer, F. (2009, April). —Social spam detection‖. In Proceedings of the 5th International Workshop on Adversarial Information Retrieval on the Web (pp. 41-48)

[8] Rada Mihalcea , Carlo Strapparava, The lie detector: explorations in the automatic recognition of deceptive language, Proceedings of the ACL-IJCNLP.