IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Enhancing Fake News Detection System Using Python & Machine Learning

Pranay Arvind Jadhav
Department of Computer, GMVIT ,Tala Raigad University of Mumbai

Omkar Santosh Kadam
Department of Computer, GMVIT, Tala Raigad University of Mumbai

Rohit Santosh Warge
Department of Computer, GMVIT, Tala Raigad University of Mumbai

Shailesh Shashikant Gothal
Department of Computer, GMVIT, Tala Raigad University of Mumbai

Prof.Sanjay Gawali Assistant Professor
Department of Computer, GMVIT tala Raigad University of Mumbai

Abstract: Recent political events have lead to an increase in the popularity and spread of fake news. As demonstrated by the widespread effects of the large onset of fake news, humans are inconsistent if not outright poor detectors of fake news. With this, efforts have been made to automate the process of fake news detection. The most popular of such attempts include “blacklists" of sources and authors that are unreliable. While these tools are useful, in order to create a more complete end to end solution, we need to account for more difficult cases where reliable sources and authors release fake news. As such, the goal of this project was to create a tool for detecting the language patterns that characterize fake and real news through the use of machine learning and natural language processing techniques. The results of this project demonstrate the ability for machine learning to be useful in this task. We have built a model that catches many intuitive indications of real and fake news as well as an application that aids in the visualization of the classification decision.

Index Terms **– OpenCV, NumPy, MySQL, Visual Studio, Machin Learnig.**

1. INTRODUCTION

Fake news exist way before from social media but it multifold when social media was introduced. Fake news is a news designed to deliberately spread hoaxes, propaganda and disinformation. Fake news stories usually spread through social media sites like Facebook, Twitter etc. As an increasing amount of our lives is spent interacting online through social media platforms, more and more people tend to seek out and consume news from social media rather than traditional news organizations. The reasons for this change in consumption behaviors are inherent in the nature of these social media platforms: (i) it is often more timely and less expensive to consume news on social media compared with traditional news media, such as newspapers or television; and (ii) it is easier to further share, comment on, and discuss the news with friends or other readers on social media. For example, 62 percent of U.S. adults get news on social media in 2016, while in 2012, only 49 percent reported seeing news on social media1. It was also found that social media now outperforms television as the major news source2. Despite the advantages provided by social media, the quality of news on social media is lower than traditional news organizations. However, because it is cheap to provide news online and much faster and easier to disseminate through social media, large volumes of fake news, i.e., those news articles with intentionally false information, are produced online for a variety of purposes, such as _nancial and political gain. It was estimated that over 1 million tweets are related to fake news \Pizzagate"3 by the end of the presidential election.

LITERATURE REVIEW

2.1 Survey of existing system

- Study1

Shloka Gilda presented concept approximately how NLP is relevant to stumble on fake information. They have used time period frequency-inverse record frequency (TF_x0002_IDF) of bi-grams and probabilistic context free grammar (PCFG) detection. They have examined their dataset over more than one class algorithms to find out the great model. They locate that TF-IDF of bi-grams fed right into a Stochastic Gradient Descent model identifies non-credible resources with an accuracy of 77.2%

- Study 2

Marco L. Della offered a paper which allows us to recognize how social networks and gadget studying (ML) strategies may be used for faux news detection .They have used novel ML fake news detection method and carried out this approach inside a Facebook Messenger chat bot and established it with a actual-world application, acquiring a fake information detection accuracy of 70.7%

- Study 3

Cody Buntain advanced a method for automating fake news detection on Twitter. They applied this method to Twitter content sourced from Buzz Feed's fake news dataset. Furthermore, leveraging non-professional, crowd sourced people instead of journalists presents a beneficial and much less costly way to classify proper and fake memories on Twitter rapidly.

- Study 4

Saranya Krishnan used superior framework to indentify faux information contents. Initially, they've extracted content material capabilities and consumer functions via Twitter API. Then functions together with statistical analysis of twitter user accounts, reverse picture searching, verification of fake news assets are used by facts mining algorithms for class and analysis.

2.2 Limitation on existing system

- Storage requirements problems
- Vulnerable detection
- While detecting the News, accuracy or time consuming problem.

3. PROBLEM STATEMENT

About detecting fake news with Python. This advanced python project of detecting fake news deals with fake and real news. Using sklearn, we build a Tfidfvectorizer on our dataset. Then, we initialize a PassiveAggressive Classifier and fit the model. In the end, the accuracy score and the confusion matrix tell us how well our model fares.

4. PROJECT OBJECTIVES

The main objective is to detect the fake news, which is a classic text classification problem with a straight forward proposition. It is needed to build a model that can differentiate between "Real" news and "Fake" news.

The goal of this project is to find the effectiveness and limitations of language-based techniques for detection of fake news through the use of machine learning algorithm including but not limited to convolutional neural networks and recurrent neural networks.

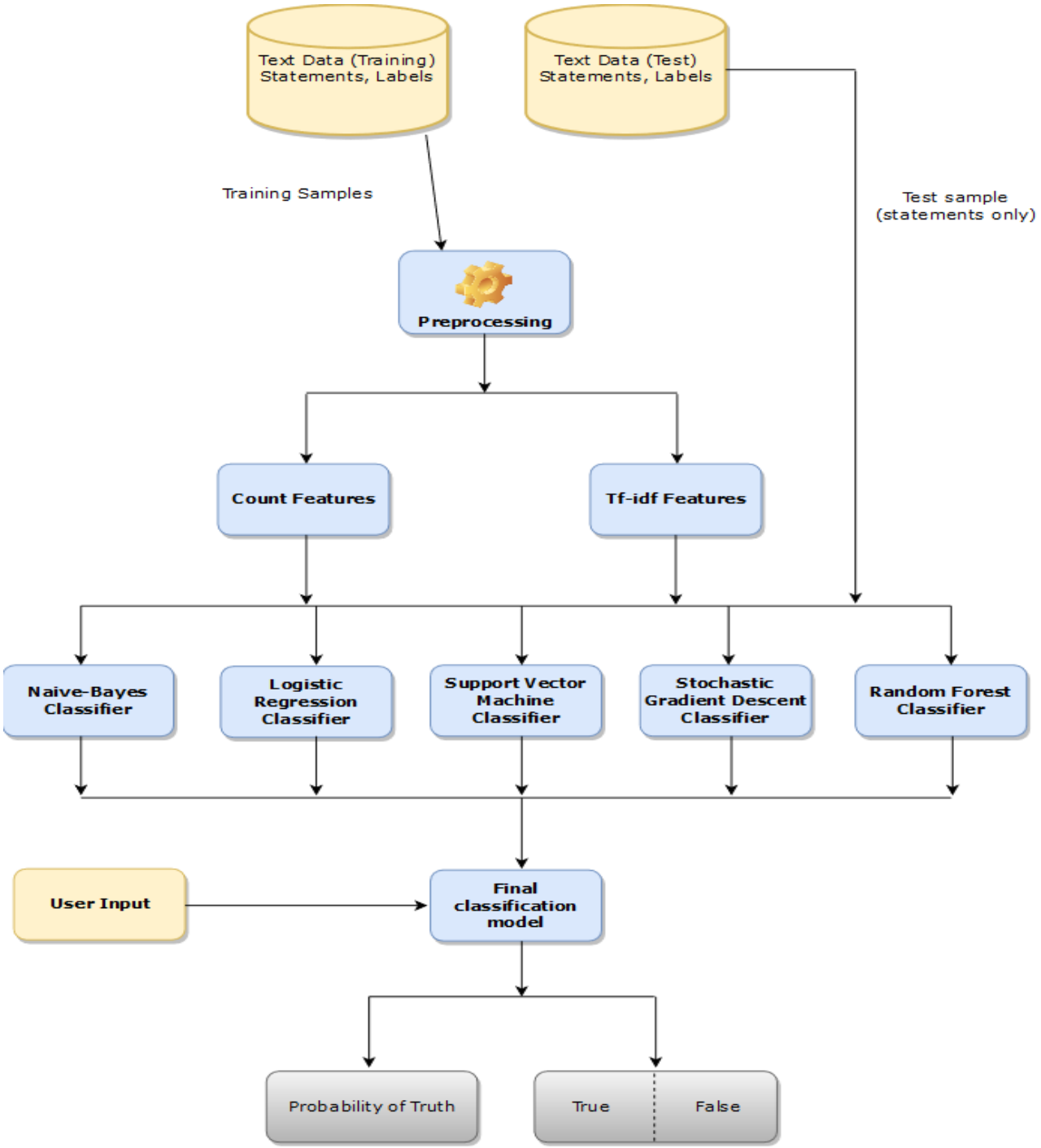
3. PROPSED SYSTEM

3.1 Algorithm

We have implementing our project work using a Python . Open source libraries of python like NumPY.

- This project aims to develop a method for detecting and classifying the news stories using natural language processing.
- The main goal is to identify fake news, which is a classic text classification issue.
- We gathered our data, pre-processed the text, and translated our article into supervised model features.
- Our goal is to develop a model that classifies a given news article as either fake or true.

.Flow and Processing ofAlgorithm :-

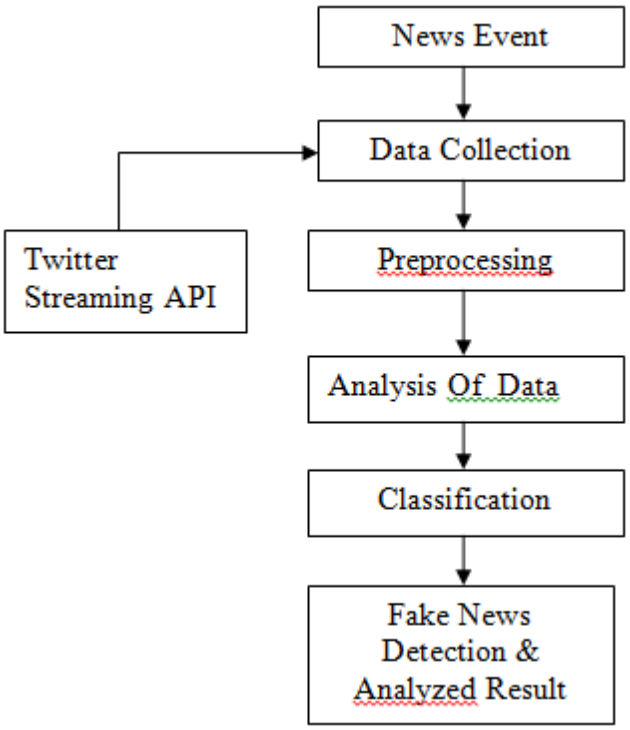


Software Requirements:

- Operating System : Windows
- Front End : Python, jupyter.

3.2 Methodology

The basic idea of our project is to build a model that can predict the credibility of real time news events. As shown in Fig., the proposed framework consists of four major steps: Data collection, Data preprocessing, Classification and Analysis of results.
We first take key phrases of the news event as an input that the individual need to authenticate. After that live data is collected from Twitter Streaming API. The filtered data is stored in the database (Mongo DB). The data preprocessing unit is responsible for preparing a data for further processing. Classification will be based on various news features, twitter reviews like Sentiment Score ,Number of Tweets ,Number of followers ,Number of hashtags ,is verified User ,Number of retweets and NLP techniques.
We are going to describe fake news detection method based on one artificial intelligence algorithm –Naïve Bayes Classifier. Sentiment Score will be calculated using Text Vectorization algorithm and NLTK(Natural Language Toolkit). By doing the evaluation of effects acquired from classification and analysis, we are able to decide the share of news being fake or real.



1. REVIEW CONCLUSION

Many people consume news from social media instead of traditional news media. However, social media has also been used to spread fake news, which has negative impacts on individual people and society. In this paper, an innovative model for fake news detection using machine learning algorithms has been presented. This model takes news events as an input and based on twitter reviews and classification algorithms it predicts the percentage of news being fake or real.

REFERENCES

[1] M. Granik and V. Mesyura, "Fake news detection using naive Bayes classifier," 2017 IEEE 1st Ukr. Conf. Electr. Comput. Eng. UKRCON 2017 - Proc., pp. 900–903, 2017.
[2] P. R. Humanante-Ramos, F. J. Garcia-Penalvo, and M. A. Conde-Gonzalez, "PLEs in Mobile Contexts: New Ways to Personalize Learning," Rev. Iberoam. Tecnol. del Aprendiz., vol. 11, no. 4, pp. 220–226, 2016.
[3] M. Risdal. (2016, Nov) Getting real about fake news. [Online]. Available: https://www.kaggle.com/ mrisdal/fake-news
[4] J. Soll, T. Rosenstiel, A. D. Miller, R. Sokolsky, and J. Shafer. (2016, Dec) The long and brutal history of fake news. [Online]. Available: https://www.politico.com/magazine/story/2016/12/ fake-news-history-long-violent-214535