Real and Fake News Detection Smart System Using Passive Aggressive Algorithm (Supervised Machine Learning)

Vaishnavi Kesharwani[1], Vaishnavi Ladole[2], Shraddha Tak[3], Vaishnavi Gaigol[4],
Prof. V. B. Bhagat[5], Dr. V. R. Thakare[6]

Students, Department of Computer Science and Engineering[1,2,3,4]
Assistant Professor, Department of Computer Science and Engineering[5]
Professor, P.G. Department of Computer Science and Engineering[6]
P. R. Pote Patil College of Engineering and Management, Amravati, Maharashtra[1,2,3,4,5]
Sant Gadge Baba Amravati University, Amravati, Maharashtra, India[6]

Abstract: *Today where the internet is ubiquitous, everyone intake news from various online platforms . 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 spam of time. The spread of fake news has far-reaching consequences like the creation of biased opinions to swaying election outcomes for the benefit of certain candidates. Moreover, spammers use appealing news headlines to generate revenue using advertisements via clickbaits. 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(Supervised ML). Our aim is to provide the news is authentic or fake.*

Keywords: Machine Learning, Raspberry Pi

I. INTRODUCTION

As an increasing amount of people spent interacting online through social media platforms, more and more people tends to hunt out and consume news from social media instead of traditional news organizations.[1] The explanations for this alteration in consumption behaviours are inherent within the nature of those social media platforms: it's often more timely and fewer expensive to consume news on social media compared with traditional journalism , like newspapers or television. 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. However, because it's inexpensive to supply news online and far faster and easier to propagate through social media, large volumes of faux news, i.e., those news articles with intentionally false information, are produced online for a spread of purposes, like financial and political gain. it had been estimated that over 1 million tweets are associated with fake news by the top of the presidential election. Given the prevalence of this new phenomenon, "Fake news" was even named the word of the year by the Macquarie dictionary in 2016.

The extensive spread of false news can have a significant negative impact on individuals and society. First, fake news can shatter the authenticity equilibrium of the news ecosystem for instance; it's evident that the most popular fake news was even more outspread on Facebook than the most accepted genuine mainstream news during the U.S. 2016 presidential election. Second, fake news intentionally persuades consumers to simply accept biased or false beliefs. Fake news is typically manipulated by propagandists to convey political messages or influence for instance, some report shows that Russia has created fake accounts and social bots to spread false stories. Third, fake news changes the way people interpret and answer real news, for instance, some fake news was just created to trigger people's distrust and make them confused; impeding their abilities to differentiate what's true from what's not. To assist mitigate the negative effects caused by fake news (both to profit the general public and therefore the news ecosystem). It's crucial that we build up methods to automatically detect fake news broadcast on social media [3]. Internet and social media have made the access to the news information much easier and comfortable [2]. Often Internet users can pursue the events of their concern in online form, and

increased number of the mobile devices makes this process even easier. But with great possibilities come great challenges. Mass media have an enormous influence on the society, and because it often happens, there's someone who wants to require advantage of this fact. Sometimes to realize some goals mass-media may manipulate the knowledge in several ways. This result in producing of the news articles that isn't completely true or maybe completely false. There even exist many websites that produce fake news almost exclusively

1.1 Aim and Objectives
- To be aware of fake articles, news while forwarding to others
- To reveal True stories
- To prevent from false crisis events
- To be Informative

1.2 Motivation
Machine learning (ML) is a type of Artificial Intelligence (AI) that allows software applications to become more accurate at predicting outcomes without being explicitly programmed to do so. Machine learning algorithms use historical data as input to predict new output values. The extensive spread of false news can have a significant negative impact on individuals and society. First, fake news can shatter the authenticity equilibrium of the news ecosystem for instance. Understanding the truth of news and message with news detection can create positive impact on the society.

1.3 Scope
- Massive amounts of data gave birth to AI systems that are already producing human-like synthetic texts, powering a new scale of disinformation operation. Based on Natural Language Processing (NLP) techniques, several lifelike text- generating systems have proliferated and they are becoming smarter every day.
- Even though TF-IDF classifiers worked well, there are possibilities of exploring other features to improve the model and make it a generic fit.
- While the project focused on text-based news articles and language models, AI algorithms can also analyze other features such as images, videos, date and time, sources, website, and domain for valuable information.

III. LITERATURE REVIEW

3.1 History
After the invention of the printing press in the 15th century, news (both real and fake) was able to spread faster than ever before. This technology meant books and other documents could be produced much quicker than any handwriting. In the mid-1700s, the printing press helped to spread fake news about George II, who was the King of Great Britain and Ireland at the time. The King was facing a rebellion, and relied on being seen as a strong leader to make sure the rebellion didn't succeed. Fake news about the King being was ill printed from sources on the side of the rebels. It didn't take long before these stories were seen by other printers who then republished them. This harmed the King's public image, and although the rebellion wasn't successful, showed how fake news can be used to try and change people's opinions.The same thing happens today when a fake story is published on purpose to harm someone else. Unfortunately, if it isn't fact-checked and gets shared by people or organisations thinking it is true, some of whom might have large followings, it starts to be taken seriously and the more it is shared the quicker it spreads… unchecked.

3.2 Related Work
[1] Himank Gupta et. al. gave a framework based on different machine learning approach that deals with various problems including accuracy shortage, time lag and high processing time to handle thousands of tweets in 1 sec. Firstly, they have collected 400,000 tweets from HSpam14 dataset. Then they further characterize the 150,000 spam tweets and 250,000 non-spam tweets. They also derived some lightweight features along with the Top-30 words that are providing highest information gain from Bag-ofWords model. 4. They were able to achieve an accuracy of 91.65% and surpassed the existing solution by approximately18%.

[2] Mykhailo Granik et. al. in their paper shows a simple approach for fake news detection using naive Bayes classifier. This approach was implemented as a software system and tested against a data set of Facebook news posts. They were collected from three large Facebook pages each from the right and from the left, as well as three large mainstream political news pages (Politico, CNN, ABC News). They achieved classification accuracy of approximately 74%. Classification accuracy for fake news is slightly worse. This may be caused by the skewness of the dataset: only 4.9% of it is fake news.

[3] Shivam B. Parikh et. al. aims to present an insight of characterization of news story in the modern diaspora combined with the differential content types of news story and its impact on readers. Subsequently, we dive into existing fake news detection approaches that are heavily based on textbased analysis, and also describe popular fake news datasets. We conclude the paper by identifying 4 key open research challenges that can guide future research. It is a theoretical Approach which gives Illustrations of fake news detection by analysing the psychological factors.

3.3 Summary

We learned to detect fake news with Python. We took a news dataset from 1900 to 2022 , implemented a TfidfVectorizer, initialized a Passive Aggressive Classifier, and fit our model. We ended up obtaining an accuracy of 75% in magnitude.

IV. DESIGN AND IMPLEMENTATION

4.1 Proposed System

The system is a Web application which help user to detect the fake news. We have given the text box where the user has the option to paste the message. All the user gives data to detector may save for further use in order to update the statue of model, data analysis in future.

4.2 System Design
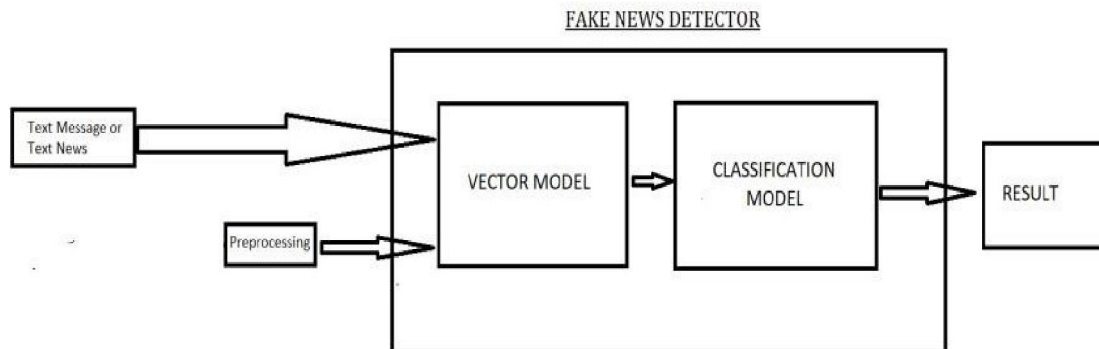
Confusion matrix: Basically this metrics how many results are correctly predicted and how many results are not correctly predicted.

Fig3: Model of confusion matrix

4.3 Experimental Setup

A. Hardware Requirement

- System: i3 & more 2.4 GHz Processor
- Hard Disk: 500 GB
- Ram: 4 GB

B. Software Requirement

- Operating System: Windows 10

- Technology Used: Python
- IDE: Pycharm

C. Implementation Steps

1. Download Dataset from Kaggle website.
2. Libraries required for the project.
- **numpy:** NumPy is a library for the Python programming language, adding support for large, multi-dimensional arrays and matrices, along with a large collection of high-level mathematical functions to operate on these arrays.
- **pandas:** Pandas is a software library written for the Python programming language for data manipulation and analysis. In particular, it offers data structures and operations for manipulating numerical tables and time series.
- **sklearn:** Scikit-learn is a free software machine learning library for the Python programming language. It features various classification, regression and clustering algorithms.
- **nltk:** The Natural Language Toolkit, or more commonly NLTK, is a suite of libraries and programs for symbolic and statistical natural language processing for English written in the Python programming language.
3. Remove all non words from news column
4. Remove all stop words using nltk
5. Then create features using TfidfVectorizer
6. Then create X as input variable having feature and Y as output variable having label 0 for fake and 1 for real.
7. Then split data in training and testing.
8. Then pass training data to Passive Aggressive Classifier algorithm. Passive Aggressive Classifier belongs to the category of online learning algorithms in machine learning. It works by responding as passive for correct classifications and responding as aggressive for any miscalculation.
- Passive: If the prediction is correct, keep the model and do not make any changes. i.e., the data in the example is not enough to cause any changes in the model.
- Aggressive: If the prediction is incorrect, make changes to the model. i.e., some change to the model may correct it.
9. Then we will check accuracy of model and deploy it on flask web framework.
10. Our web application will take a news as input and will classify it as Fake or Real according to model.

V. 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. The best model, i.e. the model with highest accuracy is used to classify the news headlines or articles. As evident above for static search, 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. We intent to build our own dataset which will be kept up to date according to the latest news. All the live news and latest data will be kept in a database using Web Crawler and online database.

REFERENCES

[1]. H. Gupta, M. S. Jamal, S. Madisetty and M. S. Desarkar, "A framework for real-time spam detection in Twitter," 2018 10th International Conference on Communication Systems & Networks (COMSNETS), Bengaluru, [2018, pp. 380-383]

[2]. 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.

[3]. Shivam B. Parikh et. al. aims to present an insight of characterization of news story in the modern diaspora combined with the differential content types of news story . [Online]. [Feb. 6, 2017]

[4]. Fake news websites. (n.d.) Wikipedia. [Online]. Available: https://en.wikipedia.org/wiki/Fake_news_website. Accessed [Feb. 05, 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 Pro

[7]. 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]

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

[9]. Kushal Agarwalla, Shubham Nandan, Varun Anil Nair, D. Deva Hema, ―Fake News Detection using Machine Learning and Natural Language Processing,‖ International Journal of Recent Technology and Engineering (IJRTE) ISSN: 2277-3878, Volume-7, Issue-6, March 2019

[10]. H. Gupta, M. S. Jamal, S. Madisetty and M. S. Desarkar, "A framework for real-time spam detection in Twitter," 2018 10th International Conference on Communication Systems & Networks (COMSNETS), Bengaluru, 2018, pp. 380-383