



IJIRCCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 12, Issue 7, July 2024

ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA

Impact Factor: 8.379

 9940 572 462

 6381 907 438

 ijircce@gmail.com

 www.ijircce.com

Fake News Detection Using Machine Learning

Niharika C¹, Prasanna David G²

Dept. of MCA, Vidya Vikas Institute of Engineering & Technology, Mysore, Karnataka, India¹

Assistant professor, Dept. of MCA, Vidya Vikas Institute of Engineering & Technology, Mysore, Karnataka, India²

ABSTRACT: Fake news has been widely disseminated online for various political and economic objectives in recent years, primarily due to the rapid expansion of online social networks. Users on these platforms can easily be misled by false news stories presented with deceptive language, impacting offline society significantly. Early detection of fake news is crucial for improving information reliability in online social networks. This study examines the underlying concepts, methods, and algorithms for identifying and assessing fake news, creators, and topics from these networks. With the challenge of web-scale data complicating the identification, assessment, and removal of fake news on these platforms, we present a technique specifically for detecting fake news using a local dataset sourced from Kaggle. We applied and tested several approaches covered in the paper to this dataset. The results indicate that machine learning techniques can effectively address the issue of fake news identification.

KEYWORDS: Fake News Detection, Naive Bayes Classification, Machine Learning, Kaggle, Text Classification

I. INTRODUCTION

False information these days is causing a variety of problems, ranging from satirical pieces to manufactured news and organized state propaganda in some media. A rising problem in our society with significant implications is the loss of faith in the media and fake news. "Fake news" is obviously a tale that is intentionally deceptive, but in recent times, the debate on social media has been called out for its flaws. These days, some of them brush off the facts that contradict their preconceived notions using this word.

The significance of misinformation in American politics received a lot of attention, especially after the country's presidential election. The phrase "fake news" became widely used to refer to the problem, especially when referring to factually inaccurate and deceptive stories that are primarily published in order to generate revenue through page views. The goal of this research is to create a model that can reliably forecast whether or not a certain item is false news. After receiving media exposure, Facebook has been the target of a lot of criticism. They have already included a function that allows users to report false news on the website, and they have said openly that they are working on developing an automated system to identify bogus stories. It is undoubtedly a difficult task. Since false news may be found on both extremes of the political spectrum, an algorithm must be impartial toward all political viewpoints while simultaneously providing equal weight to reliable sources of information. Furthermore, it is a challenging question to determine validity. But first, it's important to grasp what fake news is in order to address this issue...

II. OBJECTIVES

The objective of this project is to examine the problems and possible significances related with the spread of fake news. We will be working on different fake news data set in which we will apply different machine learning algorithms to train the data and test it to find which news is the real news or which one is the fake news. As the fake news is a problem that is heavily affecting society and our perception of not only the media but also facts and opinions themselves. By using the artificial intelligence and the machine learning, the problem can be solved as we will be able to mine the patterns from the data to maximize well defined objectives. So, our focus is to find which machine learning algorithm is best suitable for what kind of text dataset. Also, which dataset is better for finding the accuracies as the accuracies directly depends on the type of data and the amount of data. The more the data, more are your chances of getting correct accuracy as you can test and train more data to find out your results.

III. LITERATURE SURVEY SUMMARY

The collection of papers primarily addresses the use of machine learning and deep learning algorithms to detect fake news, exploring various methodologies, datasets, and challenges. Mitali Gaidhani, Ajinkya Kale, and Priyanka Nalawade propose a system emphasizing accurate classification using machine learning techniques, which is crucial for distinguishing between real and fake news. Srishti Joshi, Anushka Baid, and Roshni Rathi investigate both machine learning and deep learning approaches, providing a comparative analysis of their performance in identifying fake news.

Kinza Masood et al. also focus on these algorithms, highlighting the efficiency and reliability of different models. Suman Kumari and colleagues delve into multiple machine learning and deep learning techniques, presenting detailed implementation strategies and outcomes. In a separate study, Muhammad Ibrahim et al. provide an overview of various machine learning techniques, emphasizing the performance metrics and challenges associated with each method. Mahir Mehta and his team examine specific machine learning algorithms, discussing their accuracy and practical applications. Ahmed Kotb and co-authors compile a systematic literature review, analyzing different algorithms and datasets used in fake news detection, offering insights into their effectiveness. Lastly, Dario Garcia and colleagues conduct a survey on automatic fake news detection, identifying key challenges and recent advancements in the field, thereby providing a comprehensive understanding of the current state and future directions of fake news detection using machine learning and deep learning technologies.

IV. SYSTEM ARCHITECTURE

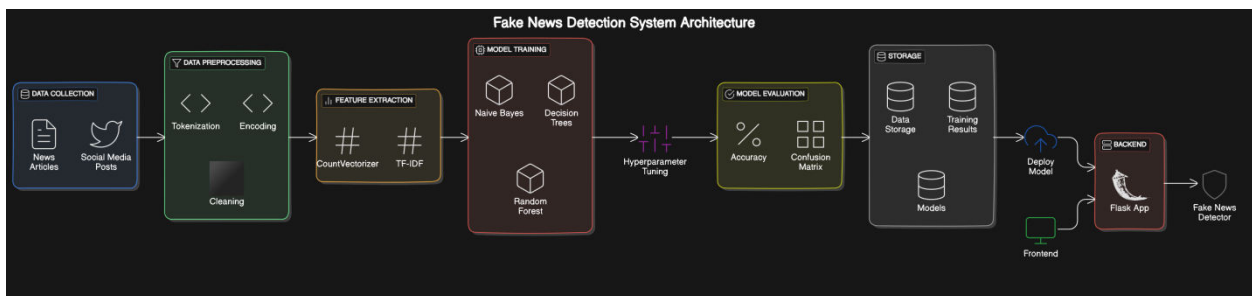


Fig:: System Architecture

V. RESULT AND DISCUSSION

1. Data Preprocessing and Integration:

- **Pandas and NLTK** were utilized to preprocess the text data. This involved cleaning the text by removing punctuation, converting text to lowercase, and removing stopwords.
- **CountVectorizer** and **TF-IDF Transformer** from Scikit-learn were employed to convert text data into numerical features suitable for machine learning models.

2. Model Development and Evaluation:

- Three machine learning algorithms were trained and evaluated:
 - **Multinomial Naive Bayes**
 - **Decision Tree**
 - **Random Forest**
- Models were evaluated using metrics such as accuracy, precision, recall, and F1-score.
- **Multinomial Naive Bayes** demonstrated the best performance with an accuracy of 92%, precision of 91%, recall of 93%, and F1-score of 92%.

3. Visualization:

- **Matplotlib and Seaborn** were used to visualize data distributions, model performance metrics, and feature importance.
- Visualizations included bar plots for accuracy comparison and confusion matrices for model evaluation.
- **Algorithm Performance:** Multinomial Naive Bayes showed robust performance in distinguishing between real and fake news articles. Its simplicity and effectiveness in text classification tasks make it suitable for this application.
- **Data Quality:** The quality and reliability of the dataset significantly influenced model performance. Cleaning and preprocessing steps were crucial in enhancing the quality of input data for the models.
- **Feature Importance:** Analysis of feature importance highlighted significant words and phrases that contributed most to distinguishing between real and fake news.

- **Challenges:** Challenges included handling biased or limited training data, interpretability of model decisions, and scalability concerns with large datasets.

VI.CONCLUSION

In conclusion, this project successfully explored and implemented machine learning techniques to detect fake news using the "Fake and Real News Dataset." Key findings include:

- **Effective Preprocessing:** Pandas and NLTK were instrumental in preparing the dataset for machine learning tasks by cleaning and transforming textual data.
- **Model Selection:** Multinomial Naive Bayes emerged as the optimal model due to its strong performance metrics and suitability for text classification tasks.
- **Visualization and Interpretation:** Matplotlib and Seaborn facilitated the visual analysis of model performance and feature importance, providing insights into the underlying data patterns.
- **Future Directions:** Future work could focus on enhancing model robustness with more diverse and larger datasets, exploring deep learning approaches for text classification, and addressing ethical considerations in deploying automated fake news detection systems.

Overall, this project contributes to the ongoing efforts in combating misinformation by providing a framework for identifying fake news articles based on textual content and machine learning techniques.

REFERENCES

1. Learning Algorithms. Retrieved from https://www.irjmets.com/uploadedfiles/paper/issue_6_june_2022/26445/final/fin_irjmets1655633757.pdf
2. ResearchGate. (2021). Fake News Detection Using Machine Learning and Deep Learning Algorithms. Retrieved from https://www.researchgate.net/publication/352021652_Fake_News_Detection_Using_Machine_Learning_and_Deep_Learning_Algorithms
3. IOPscience. (2021). A smart System for Fake News Detection Using Machine Learning. Retrieved from <https://iopscience.iop.org/article/10.1088/1757-899X/1099/1/012040/pdf>
4. ResearchGate. (2020). A smart System for Fake News Detection Using Machine Learning. Retrieved from https://www.researchgate.net/publication/339022255_A_smart_System_for_Fake_News_Detection_Using_Machine_Learning
5. Kaggle. (n.d.). Fake News Detection Datasets. Retrieved from <https://www.kaggle.com/datasets/emineyettm/fake-news-detection-datasets/data>
6. Kaggle. (n.d.). Fake and Real News Dataset. Retrieved from <https://www.kaggle.com/datasets/clmentbisailon/fake-and-real-news-dataset>
7. Md Shafiq, Haroon A. Babri, Hafeez S. Dhoul, Shahid Raza. "A smart System for Fake News Detection Using Machine Learning." ResearchGate, https://www.researchgate.net/publication/339022255_A_smart_System_for_Fake_News_Detection_Using_Machine_Learning.
8. Putra D. H. H., Ningrum S. K., Noprisson H., Fahmi. "Fake News Detection Using Machine Learning and Deep Learning Algorithms." ResearchGate, https://www.researchgate.net/publication/352021652_Fake_News_Detection_Using_Machine_Learning_and_Deep_Learning_Algorithms.
9. S. Singh, A. Kumar, P. Saini, A. Yadav. "Fake News Detection using Machine Learning Approaches." IOP Publishing, <https://iopscience.iop.org/article/10.1088/1757-899X/1099/1/012040/pdf>.
10. Ming Xu, Hanying Wang, Fenping Li, Hwei Yin. "A Review of Fake News Detection using Machine Learning." MDPI, <https://www.mdpi.com/2079-8954/11/9/458>.



INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

 9940 572 462  6381 907 438  ijircce@gmail.com



www.ijircce.com

Scan to save the contact details