iJETRM
International Journal of Engineering Technology Research & Management
Published By:
https://www.ijetrm.com/

FAKE NEWS DETECTION USING MACHINE LEARNING

Hemalatha V
Assistant Professor, Department of Computer Science and Engineering, N.S.N College of Engineering and Technology, Tamilnadu, India.

Dinesh M
PG Student, Department of Computer Science and Engineering, N.S.N College of Engineering and Technology, Tamilnadu, India.

ABSTRACT
The rampant spread of false news via online platforms has been a pressing issue over the past few years. With the speedy expansion of social media and online news reading, the capacity to authenticate the truthfulness of news has been more crucial than ever. This paper introduces a Fake News Detection System based on machine learning algorithms for categorizing news articles as true or false. The system utilizes various preprocessing methods, TF-IDF feature extraction, and classification algorithms like Logistic Regression, Decision Tree, Random Forest, and Gradient Boosting. A special result analysis and visualization module provides insight into system performance. The architecture is modular, scalable, and easily retrainable with new data. Results indicate positive accuracy and real-world applicability for online platforms.

Keywords:
Fake News Detection, Machine Learning, Natural Language Processing, Text Classification, TF-IDF, Visualization

INTRODUCTION

Fake news is intentionally misleading information, usually disseminated to mislead readers or influence public opinion. With the internet and social media sites becoming increasingly prominent, such information spreads quickly. It is impossible for manual fact-checking to keep pace with the speed and quantity of content creation. This requires automated systems capable of identifying fake news through computational means. We suggest in this paper a machine learning-based method that classifies fake news articles from textual features. Our system consists of modules for input processing, classification, analysis, and visualization.

OBJECTIVES

This research proposes an approach to create a machine learning-based Fake News Detection System that identifies news articles as fake or authentic. It constructs an end-to-end pipeline, ranging from data acquisition and cleaning to feature extraction and classification, ensuring the system is accurate and scalable. Through model comparison, the system chooses the optimal algorithm for production while focusing on user experience through clear predictions and visualization metrics. Finally, it presents a modular, real-time solution against spreading misinformation over the internet.

METHODOLOGY

The approach taken by this project is a systematic, step-by-step methodology that makes it possible to detect false news using machine learning methods. The first step is data acquisition, where a labeled dataset of both fake and actual news articles is collected from available public sources. This dataset acts as the basis for training and testing models.

Once the data is gathered, it goes through a comprehensive preprocessing phase. This includes text cleaning where all letters are made lower case, and punctuation, special characters, digits, and stopwords are all excluded. Tokenization is performed after that, and the step helps in standardizing the input text and removing noise, which can be easily interpreted by the machine learning models.

iJETRM
International Journal of Engineering Technology Research & Management
Published By:
https://www.ijetrm.com/

After preprocessing, the second phase is feature extraction. For this system, the Term Frequency–Inverse Document Frequency (TF-IDF) method is employed to transform the preprocessed text into numerical feature vectors. TF-IDF gives weights to words based on their occurrence and importance in the whole dataset, retaining the specificity of each word regarding the detection of fake news.

The preprocessed data is then fed into the machine learning models. The system is trained with a range of algorithms, such as Logistic Regression, Decision Tree, Random Forest, and Gradient Boosting. Each model is separately trained and tested based on performance metrics like accuracy, precision, recall, and F1 score. The model with the highest performance is then chosen for deployment.

Following prediction, the output is represented graphically with the aid of graphs and charts that give information on the performance of the model. A result analysis and visualization module presents outputs like confusion matrices and ROC curves, which enable users to get a sense of the model's strengths and weaknesses. The system architecture also accommodates a feedback system for retraining with new data in order to enable long-term adaptation and enhancement.

The whole system has modular architecture, which facilitates smooth integration of every step from data entry to the visualization of results. The backend uses Flask, while the user interface provides real-time interaction with the model, and hence the system is feasible for daily utilization.

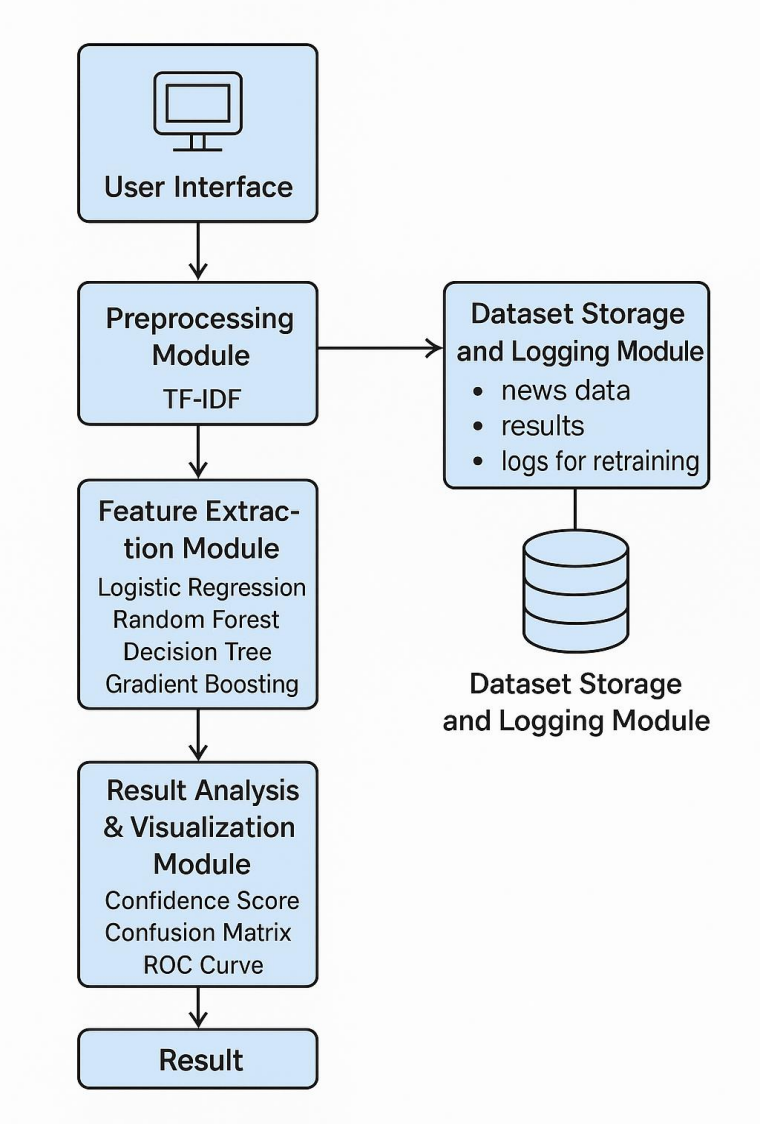

Figure 1 Fake news detection flowchart

iJETRM
International Journal of Engineering Technology Research & Management
Published By:
https://www.ijetrm.com/

RESULTS AND DISCUSSION

The Fake News Detection System was tested extensively with a labeled dataset that contained real and fake news articles retrieved from publicly accessible repositories. The data was preprocessed and partitioned into training and test sets in order to measure the performances of different machine learning models deployed in the system. The major models experimented are Logistic Regression, Decision Tree, Random Forest, and Gradient Boosting.

Of the models, Random Forest and Gradient Boosting exhibited better performance with regards to classification accuracy, both with accuracy greater than 90% on the test data. Logistic Regression, although less accurate, was faster and more interpretable and thus more ideal for real-time use. Decision Tree models, although easy to visualize and interpret, were more likely to overfit than the ensemble models.

Besides accuracy, other metrics like precision, recall, and F1-score were also used to evaluate the reliability of the model. Random Forest had the best F1-score value, which reflects a well-balanced precision and recall. The confusion matrices for all models were obtained for detailed insight into true positives, true negatives, false positives, and false negatives, and they showed that the majority of the misclassifications were done in borderline or vague news articles.

The system further consists of a result visualization module that plots ROC curves and precision-recall curves. The plots assisted in examining the trade-off between various thresholds and assessing how well the models can differentiate between fake and real news. A high area under the ROC curve (AUC > 0.95) ensured strong discriminatory power by the models.

ACKNOWLEDGEMENT

I genuinely appreciate everyone who helped me to finish this project. I am thankful to my project supervisor, Mr. Stanley john, for their support and direction. I would also like to thank the Department of Computer Science and Engineering of N.S.N College of Engineering and Technology for their educational support.

My special thanks to my friends and teammates for pursuing this project as a team and motivating one another. I also appreciate the open-source community for their datasets and libraries. Finally, thanks to my family, for their unconditional support and faith throughout this project.

CONCLUSION

The Fake News Detection System offers a practical approach to combating misinformation using machine learning. By integrating text preprocessing, TF-IDF-based feature extraction, and multiple classification algorithms, it effectively identifies fake and real news. Ensemble models like Random Forest and Gradient Boosting enhance accuracy and reliability.

Its modular design supports easy updates, retraining, and future integration of deep learning and real-time data streaming. Visualization tools improve interpretability through metrics like confusion matrices and ROC curves.

This system lays a strong foundation for automated fake news detection, with potential expansion into real-time feeds, multi-language support, and user feedback mechanisms. As digital misinformation grows, such tools are crucial for promoting truth and transparency in online information sharing.

REFERENCES

[1] H. Jin, M. Cao, Y. Guo, and Y. Zhang, “Multimodal Fusion With Graph Convolutional Network for Fake News Detection,” *IEEE Transactions on Computational Social Systems*, vol. 8, no. 4, pp. 892–901, Aug. 2021.

[2] M. Al-Rakhami and A. Al-Amri, “Lies Kill, Facts Save: Detecting Fake News in the COVID-19 Era,” *IEEE Access*, vol. 9, pp. 155507–155541, 2021.

[3] X. Zhou and R. Zafarani, “A Survey of Fake News: Fundamental Theories, Detection Methods, and Opportunities,” *ACM Computing Surveys (CSUR)*, vol. 53, no. 5, pp. 1–40, Sep. 2020.

[4] H. Ahmed, I. Traore, and S. Saad, “Detecting opinion spams and fake news using text classification,” *Security and Privacy*, vol. 1, no. 1, pp. e9, Mar. 2018.

[5] K. Shu, A. Sliva, S. Wang, J. Tang, and H. Liu, “Fake News Detection on Social Media: A Data Mining Perspective,” *ACM SIGKDD Explorations Newsletter*, vol. 19, no. 1, pp. 22–36, Jun. 2017.