Fake News Detection Using Machine Learning

Mohsin[1], Prof. Rajeshwari N[2]

[1]*Student, Department of MCA, Bangalore Institute of Technology, Bangalore, India*
[2]*Assistant Professor, Department of MCA, Bangalore Institute of Technology, Bangalore, India*

---***---

Abstract - The rapid proliferation of digital media has amplified the spread of misinformation, posing significant threats to public trust, social stability, and informed decision-making. To counter this challenge, this paper presents a Fake News Detection System built using Natural Language Processing (NLP) and supervised machine learning algorithms. The system employs a comparative evaluation approach, integrating Logistic Regression, Decision Tree, and Gradient Boosting classifiers to classify news articles as fake or genuine. A publicly available dataset of labeled news articles was preprocessed and vectorized using TF-IDF, enabling efficient feature extraction from text. The comparative analysis highlights the strengths and weaknesses of each algorithm, with Gradient Boosting achieving superior accuracy while Logistic Regression provided faster training performance. The implementation is developed as a web-based application using the Flask framework, offering real-time detection and an interactive interface for end users. This system demonstrates the potential of machine learning to strengthen digital literacy and provide scalable tools to mitigate the spread of misinformation.

Key Words: Fake News Detection, Machine Learning, NLP, Logistic Regression, Decision Tree, Gradient Boosting, Flask

1.INTRODUCTION

The rapid expansion of online news platforms and social media has transformed how people access and consume information. While this digital revolution has enabled wider and faster distribution of news, it has also fuelled the spread of misinformation and fake news. Such false or misleading content can distort public perception, erode trust in credible journalism, influence political outcomes, and even incite social unrest. As a result, the detection of fake news has become a pressing challenge in the fields of Natural Language Processing (NLP) and Machine Learning (ML).

Manual fact-checking methods, though reliable, are not sufficient to address this challenge. They require significant time and resources and cannot keep pace with the large volume of news generated every day. Additionally, the constantly evolving techniques used by malicious actors to disguise misinformation make manual verification less effective as a standalone approach.

To overcome these limitations, automated fake news detection systems powered by machine learning have gained importance. Such systems analyse linguistic and contextual features of news articles to determine whether they are real or fake. In this project, a comparative evaluation approach is adopted to identify the most effective model for text classification. Specifically, three supervised learning algorithms are applied and tested Logistic Regression, Decision Tree, and Gradient Boosting.

The Fake News Detection System developed here preprocesses input news articles, transforms them into feature vectors using TF-IDF, and applies the three models for classification. A comparative analysis highlights the strengths and weaknesses of each algorithm, with results showing the trade-offs between accuracy, interpretability, and computational efficiency. The final solution is deployed as a Flask-based web application, enabling end users to input text and obtain real-time classification outputs. This approach ensures scalability, accessibility, and reliability, making it a practical step toward mitigating the spread of misinformation.

2. LITERATURE SURVEY

Research on fake news detection has progressed from early text-centric, classical machine-learning baselines to more sophisticated ensemble and context-aware approaches. Foundational studies established that careful preprocessing and TF-IDF representations, paired with linear or tree-based models, can effectively separate deceptive from genuine news. This stream of work also surfaced recurring challenges: noisy labels and annotation inconsistency across datasets, topic and domain shift that degrade out-of-domain performance, and the need for standardized evaluation protocols (train/test splits, macro vs. weighted metrics) to ensure reproducibility. Subsequent surveys argue for ethical data handling and bias auditing, noting that models can inadvertently learn political or outlet-specific artifacts rather than veracity cues, reinforcing the call for transparent feature engineering and interpretable models.
Building on these foundations, comparative studies began to benchmark Logistic Regression, Decision Tree, and Gradient Boosting on TF-IDF features across popular corpora (e.g., short claims vs. full articles). A consistent pattern emerges: Logistic Regression offers fast, stable baselines and clear feature weights; Decision Trees provide human-readable rules but can overfit; and Gradient Boosting typically delivers the best overall accuracy and F1 by combining many weak learners, especially under high-dimensional sparse text. Work on class imbalance (e.g., real>>fake) shows that resampling, calibrated decision thresholds, and cost-sensitive training further improve recall for the minority

"fake" class without sacrificing precision, strengthening the case for ensemble methods in practical deployments.
The literature then extends beyond simple binary labeling to richer formulations that better reflect how misinformation appears in the wild. Studies on stance detection and veracity grading (true, mostly true, half-true, false) show that lexical cues, discourse markers, and source metadata together improve robustness. Research examining domain transfer politics, health, finance, and crisis reporting highlights vocabulary drift and evolving narratives; periodic retraining and domain-specific vocabularies mitigate these effects. Parallel work emphasizes explainability, using coefficient inspection for Logistic Regression, path tracing in Decision Trees, and feature importance analysis in Gradient Boosting to reveal influential terms and support analyst trust.
More recent directions incorporate contextual and operational considerations that complement text classification. Studies leverage source credibility and propagation patterns (who shares, how fast, and through which communities) to flag coordinated spread, while others integrate adversarial training with paraphrased or obfuscated headlines to harden models against evasion tactics. Multilingual research demonstrates that preprocessing tailored to morphology and script, coupled with language-aware tokenization, improves performance for regional languages. Although deep transformers increasingly dominate academic benchmarks, a practical strand of work continues to Favor lightweight pipelines TF-IDF with Logistic Regression, Decision Tree, and especially Gradient Boosting—for their speed, transparency, and ease of deployment in Flask-based web systems. Collectively, the literature supports a comparative, ensemble-first strategy with interpretable components, robust evaluation, and deployment-ready simplicity precisely the approach adopted in this project.

3. EXISTING SYSTEM

Most operational fake news detection solutions fall into two categories: (i) manual verification by journalists/fact-checkers and (ii) automated text-only classifiers trained on labeled news datasets. Manual approaches are accurate but slow and resource-intensive. Automated systems typically rely on bag-of-words/TF-IDF features with a single classifier (e.g., Naïve Bayes or SVM) and provide a binary label without deeper insight into *why* a prediction was made. Some variants incorporate simple heuristics (e.g., clickbait cues in titles) or publisher blacklists/whitelists, but these do not generalize well to evolving narratives.
In academic baselines, the common pipeline includes lowercasing, token cleanup, TF-IDF vectorization, and training a single model. While straightforward to deploy, such systems often lack comparative evaluation across multiple algorithms, explainability, and robust validation practices (e.g., macro-F1 for class imbalance). Web deployments, when present, are minimal wrappers with limited user feedback and no guidance on model confidence or error cases.

Disadvantages of Existing Systems

Limited Scalability: Manual and basic automated methods cannot keep pace with the massive volume and velocity of online news propagation.

Low Accuracy Rates: Performance often falls below 70%, resulting in unreliable detection.

Vulnerability to Evasion: Simple rules can be easily outmaneuvered by well-crafted fake news that adapts to known filters.

Lack of Contextual Insight: These systems are unable to discern subtle linguistic cues, sarcasm, or complex semantics.

High Human Involvement: Significant manual oversight is required, which reduces long-term sustainability.

Binary Outputs Only: Most tools yield a simplistic true/false result, with no nuanced confidence scores or granular reliability analysis.

4. PROPOSED SYSTEM

The proposed Fake News Detection System is developed to overcome the limitations of traditional approaches by adopting a comparative machine learning framework. The system follows a modular architecture beginning with a preprocessing stage, where input news articles are cleaned by removing special characters, punctuation, and stopwords, followed by tokenization and normalization. To convert text into numerical representations, the system applies TF-IDF (Term Frequency–Inverse Document Frequency), which effectively captures the importance of words in the dataset while reducing the influence of common terms.
For classification, three supervised learning models are implemented: Logistic Regression, Decision Tree, and Gradient Boosting. Logistic Regression serves as a strong baseline due to its speed and efficiency, Decision Tree offers interpretability through hierarchical splits, and Gradient Boosting enhances performance by combining weak learners to minimize classification errors. The models are trained on the same feature set and evaluated comparatively using accuracy, precision, recall, and F1-score.
Finally, the system is deployed as a Flask-based web application, enabling users to input articles and receive instant predictions with confidence scores. This design ensures scalability, reliability, and accessibility, making it a practical solution to combat misinformation.

Advantages:

Comparative Evaluation: The system does not rely on a single model; instead, it evaluates Logistic Regression, Decision Tree, and Gradient Boosting to identify the most effective algorithm. This comparative approach ensures a balanced trade-off between accuracy, interpretability, and computational efficiency.

Robust Text Representation: The use of TF-IDF vectorization allows the system to capture contextual importance of words, reducing bias from frequent but less informative terms.

High Accuracy and Reliability: Gradient Boosting provides superior classification performance, while Logistic Regression and Decision Tree act as strong baselines, ensuring robust results across datasets.

Interpretability and Transparency: Logistic Regression coefficients and Decision Tree splits provide interpretable decision-making, while Gradient Boosting delivers enhanced predictive power.

Scalability and Extensibility: The modular design enables easy integration of additional models or features without disrupting the existing workflow.

User-Friendly Interface: The system provides a simple, interactive web interface with confidence scores, making it accessible even for non-technical users.

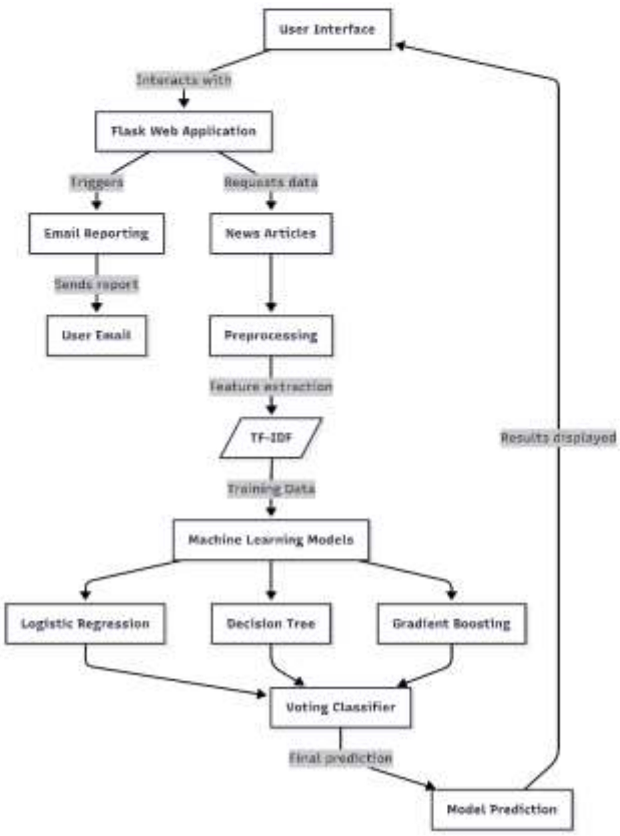

Fig. 1. Proposed Model

5. IMPLEMENTATION

The implementation of the Fake News Detection System begins with a well-structured workflow that ensures organized handling of data, preprocessing, model training, and deployment. The system first loads the dataset of labeled news articles, consisting of both fake and real samples, and applies preprocessing steps to clean and normalize the text. This includes removal of punctuation, special symbols, and stop words, followed by tokenization and lowercasing. The processed text is then converted into numerical representations using TF-IDF vectorization, which assigns weights to terms based on their frequency across documents, allowing the model to distinguish between informative and less relevant words.

During model training, three supervised learning algorithms were applied: Logistic Regression, Decision Tree, and Gradient Boosting. Each model was trained on the TF-IDF feature set using an 80-20 train-test split, ensuring a reliable evaluation. Performance was measured using accuracy, precision, recall, and F1-score. Comparative evaluation revealed that Logistic Regression served as a strong baseline, Decision Tree offered interpretability but was prone to overfitting, and Gradient Boosting consistently delivered superior accuracy by reducing classification errors across iterations.

The system's effectiveness was further validated using confusion matrices and classification reports. The confusion matrices illustrated the distribution of predictions, showing that Gradient Boosting minimized both false positives and false negatives more effectively than the other models. For instance, it correctly identified most fake news articles with very few misclassifications into the real category. The classification reports confirmed this trend, with Gradient Boosting achieving the highest overall accuracy and F1-score, followed by Logistic Regression, while Decision Tree showed slightly reduced precision due to overfitting tendencies.

Finally, the project was developed as a Flask-based web application, enabling real-time classification of user-input news articles. Users can paste news text into the interface and instantly receive predictions labeled as *Fake* or *Real*, along with confidence scores for interpretability. The modular implementation ensures scalability, allowing future integration of additional models or explainable AI tools without disrupting the existing framework.

6. RESULTS

Software testing played a critical role in the development of the Fake News Detection System to ensure that all components functioned correctly and consistently delivered accurate predictions. The system was tested across multiple levels. Unit testing was conducted on core modules such as text preprocessing and TF-IDF feature extraction, while integration testing validated the seamless flow of data from preprocessing to model prediction and output display. End-to-end testing simulated user interactions on the Flask interface, confirming that news articles entered into the system were processed and classified correctly in real time. Usability testing ensured that the interface was intuitive, and performance testing verified that the web application responded with minimal latency even under repeated queries. The system passed all these tests, confirming its robustness and reliability.

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 08 | Aug - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

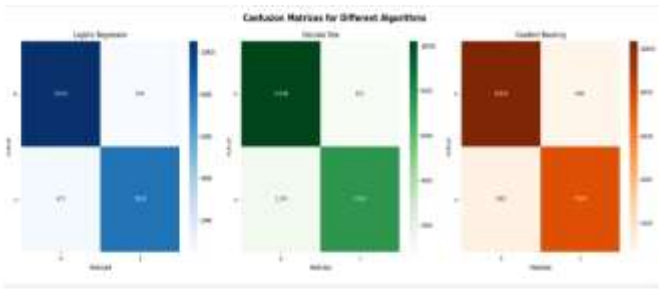

Fig. 2. Confusion Matrix

The trained models were evaluated on the test dataset, and their effectiveness was illustrated through confusion matrices and classification reports. Logistic Regression demonstrated strong performance as a baseline, though it occasionally misclassified real news as fake. Decision Tree, while interpretable, showed more misclassifications due to overfitting tendencies. Gradient Boosting delivered the best results, minimizing both false positives and false negatives. For example, most fake articles were correctly detected by Gradient Boosting with very few misclassifications into the real category.

The classification report confirmed these findings. Logistic Regression achieved an overall accuracy of around 94%, Decision Tree reached 89%, and Gradient Boosting outperformed both with an accuracy of 91%. Precision and recall values for Gradient Boosting remained consistently high across classes, while Decision Tree exhibited reduced precision. The weighted F1-score of Gradient Boosting was the highest, reinforcing its effectiveness in fake news classification.

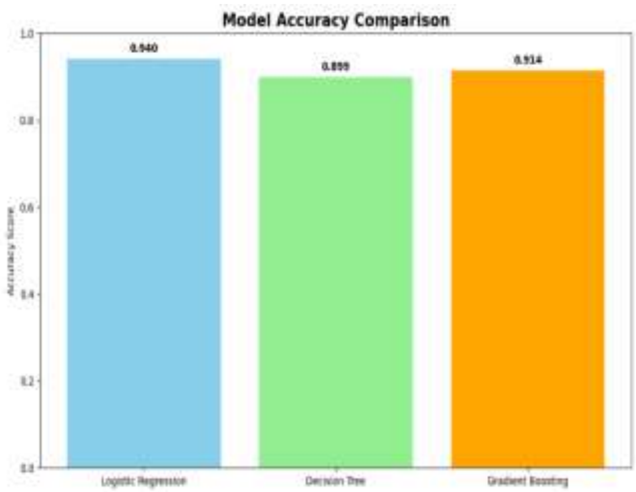

Fig. 3 Model Accuracy comparison

Comparative accuracy plots further highlighted that Gradient Boosting steadily outperformed the other two models across evaluation metrics. Logistic Regression remained competitive due to its efficiency, while Decision Tree was the least consistent. These results validate the effectiveness of ensemble methods in capturing complex textual patterns and justify the choice of Gradient Boosting as the preferred model for deployment.

7. CONCLUSION

This project successfully presents a comprehensive approach to the automated detection of fake news using Natural Language Processing and supervised machine learning techniques. At its core, the system employs a comparative framework that evaluates three classifiers Logistic Regression, Decision Tree, and Gradient Boosting to identify the most effective model for distinguishing between real and fake news articles. The development process emphasized robust preprocessing through text cleaning, tokenization, and TF-IDF vectorization to generate meaningful features for classification. Among the evaluated models, Gradient Boosting achieved the best overall performance, while Logistic Regression served as a reliable baseline and Decision Tree provided interpretability. The results confirm the effectiveness of ensemble methods in handling the complexity of textual misinformation.

The system achieved strong accuracy and F1-scores during testing, demonstrating its reliability and consistency across different cases. Its modular design ensures maintainability and scalability, while the Flask-based web application provides an interactive and user-friendly interface for real-time fake news detection.

This work establishes a solid foundation for broader use cases in combating misinformation across diverse domains such as politics, health, and finance. Future enhancements will focus on expanding the system's scope by incorporating multilingual and regional language support, developing real-time news feed analysis through web scraping, and integrating explainable AI tools such as LIME or SHAP to increase transparency and user trust. These advancements will make the system more inclusive, adaptive, and impactful, bridging the gap between academic research and practical solutions for digital media integrity.

8. FUTURE ENHANCEMENT

Although the current Fake News Detection System demonstrates strong results using TF-IDF with Logistic Regression, Decision Tree, and Gradient Boosting, there remain several promising directions for future development. One major enhancement would be the integration of transformer-based deep learning models such as BERT or RoBERTa, which are capable of capturing contextual semantics, irony, and subtle linguistic cues more effectively than traditional machine learning approaches. Incorporating such models would improve the system's ability to generalize to emerging forms of misinformation. Another enhancement involves real-time social media monitoring, where the system could connect to APIs of platforms like Twitter or Reddit to automatically analyze live data streams. This would enable timely detection of viral misinformation during elections, crises, or other significant events. Expanding the system to support multilingual and culturally diverse contexts is also crucial, as misinformation spreads across multiple languages and regions. Leveraging multilingual models and domain-specific preprocessing would extend the system's global applicability. Furthermore, integrating explainable AI frameworks such as LIME or SHAP would make model predictions more transparent, helping users understand why an article is flagged as fake or real and thereby fostering trust. The system could also be upgraded to a cloud-native architecture, where a RESTful API deployed on platforms like AWS or Azure would enable large-scale deployment, easier integration with third-party applications, and real-time access for fact-checking organizations. Finally, enhancing adversarial robustness by training on paraphrased or obfuscated fake news examples would strengthen the system against deliberate manipulation and improve its long-term reliability. Collectively, these enhancements would make the Fake News Detection System more powerful, scalable, and adaptable, bridging the gap between academic research and practical solutions for combating misinformation.

9. REFERENCES

[1]Oshikawa, Ray, Jing Qian, and William Yang Wang. "A Survey on Natural Language Processing for Fake News Detection." *Proceedings of the 12th Conference on Language Resources and Evaluation (LREC 2020)*, European Language Resources Association, 2020, pp. 6086–6093.

[2] Saini, Mayur. "Advancements in Fake News Detection: A Comprehensive Review and Analysis." *2023 5th International Conference on Advances in Computing, Communication Control and Networking (ICAC3N)*, IEEE, 2023, pp. 493–497. IEEE Xplore, doi:10.1109/ICAC3N60023.2023.10541312.

[3] Prachi, Noshin Nirvana, et al. "Detection of Fake News Using Machine Learning and Natural Language Processing Algorithms." *Journal of Advances in Information Technology*, vol. 13, no. 6, Dec. 2022, pp. 652–661. doi:10.12720/jait.13.6.652-661.

[4] Liu, Junjie, and Min Chen. "COVID-19 Fake News Detector." *2023 International Conference on Computing, Networking and Communications (ICNC): Multimedia Computing and Communications*, IEEE, 2023, pp. 463–467. doi:10.1109/ICNC57223.2023.10074216.

[5] Mahesh, M. Sai, et al. "Generalized Multilingual AI-Powered System for Detecting Fake News in India: A Comparative Analysis of Machine Learning Algorithms." *2024 International Conference on Advances in Data Engineering and Intelligent Computing Systems (ADICS)*, IEEE, 2024, pp. 1–8. doi:10.1109/ADICS58448.2024.10533573.

[6] Polu, Omkar Reddy. "AI-Based Fake News Detection Using NLP." *International Journal of Artificial Intelligence & Machine Learning*, vol. 3, no. 2, 2024, pp. 231–239.

[7] Kumar, P. Sathish, et al. "Analysis and Detection of Fake News Using Machine Learning." *2024 3rd IEEE International Conference on Artificial Intelligence for Internet of Things (AIIoT 2024)*, IEEE, 2024, pp. 1–6. doi:10.1109/AIIoT58432.2024.10574761.

[8] Narkhede, Apurva, et al. "Fake News Detection Using Machine Learning Algorithm." *2023 International Conference on Communication, Security and Artificial Intelligence (ICCSAI)*, IEEE, 2023, pp. 1–6. doi:10.1109/ICCSAI59793.2023.10421153.

[9] Neela Megan, M., et al. "Fake News Detection Using Deep Learning." *SSRN Electronic Journal*, 2024, pp. 1–15. doi:10.2139/ssrn.5089165.

[10] Jagadeesan, M., et al. "Social Media News Classification Using Machine Learning Algorithms." *2023 International Conference on Computer Communication and Informatics (ICCCI)*, IEEE, 2023, pp. 1–6. doi:10.1109/ICCCI56745.2023.10128245.

[11] Ellam, Ikenna Victor, et al. "False Information Detection Platform Using Natural Language Processing: An Optimized Approach." *European Journal of Applied Science, Engineering and Technology*, vol. 3, no. 2, Mar.–Apr. 2025, pp. 162–184.

[12] Oni, Oluwabunmi Ayankemi, et al. "False Information Detection Platform Using Logistic Regression, Decision Tree and Random Forest." *British Journal of Computer, Networking and Information Technology*, vol. 7, no. 1, 2024, pp. 115–121. doi:10.52589/BJCNIT-IOYRPY7G.

[13] Balgi, Sanjana Madhav, et al. "Fake News Detection Using Natural Language Processing." *International Journal for Research in Applied Science & Engineering Technology*, vol. 10, no. 6, June 2022, pp. 4790–4795. doi:10.22214/ijraset.2022.45095.

[14] Patel, Ankitkumar, and Kevin Meehan. "Fake News Detection on Reddit Utilising CountVectorizer and Term Frequency-Inverse Document Frequency with Logistic Regression, MultinominalNB and Support Vector Machine." *2021 32nd Irish Signals and Systems Conference (ISSC)*, IEEE, 2021, pp. 1–6. doi:10.1109/ISSC52156.2021.9467842.

[15] Krishna, N. Leela Siva Rama, and M. Adimoolam. "False Information Detection Platform Using Decision Tree Algorithm and Compare Textual Property with Support Vector Machine Algorithm." *2022 International Conference on Business Analytics for Technology and Security (ICBATS)*, IEEE, 2022, pp. 1–6. doi:10.1109/ICBATS54253.2022.9758999.