



International Journal of Innovative Research in Science Engineering and Technology (IJIRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Impact Factor: 8.699

Volume 15, Special Issue 1, January 2026

Fake News Detection

Lilesh Suralkar¹, Girish Yadav², Yashwanti Warghade³, Sanika Desale⁴, Vrushali Bhamare⁵

Department of Information Technology, Indala College of Engineering, Kalyan, Maharashtra, India^{1,2,3,4}

Professor, Department of Information Technology, Indala College of Engineering, Kalyan, Maharashtra, India⁵

ABSTRACT: The rise of social media and digital news outlets has dramatically increased the circulation of misleading information, creating significant risks to societal trust, stability, and democratic decision-making. Identifying fake news manually is difficult due to the high volume and rapid flow of online content. This study presents an automated approach for detecting fake news using advanced machine learning and natural language processing (NLP) techniques. The system processes data through preprocessing and feature extraction, followed by classification with supervised algorithms such as Naïve Bayes, Logistic Regression, and Support Vector Machines. The model's performance is evaluated using standard metrics including accuracy, precision, recall, and F1-score, confirming its ability to effectively differentiate between truthful and deceptive news articles

KEYWORDS: Automated Fake News Detection, Machine Learning Methods, NLP Techniques, News Classification

I. INTRODUCTION

The rapid growth of social media and digital news platforms has enabled the fast circulation of information, but it has also increased the spread of misleading or fake news. Fake news is intentionally false information presented as genuine news and can significantly influence public opinion, disrupt social stability, and affect democratic processes.

Traditional manual detection methods are inadequate due to the large volume and high speed of online content. Automated fake news detection using Machine Learning (ML) and Natural Language Processing (NLP) has emerged as an effective solution. This research focuses on developing a system that can accurately classify news articles as fake or real by analyzing textual data using NLP-based preprocessing and supervised ML classifiers.

II. SYSTEM MODEL AND ASSUMPTIONS

2.1 System Model

The system consists of five main modules:

1. Data Input: News articles are collected from reliable datasets containing labeled fake and real news.
2. Preprocessing Module: Raw text is cleaned, tokenized, and normalized using NLP techniques such as stop-word removal, stemming, and lemmatization.
3. Feature Extraction Module: Text is converted into numerical vectors using TF-IDF, capturing the significance of each term in the dataset.
4. Classification Module: Supervised ML algorithms such as Naïve Bayes, Logistic Regression, and SVM classify news as fake or real.
5. Output: The system displays the classification result along with the probability score for authenticity.

2.2 Assumptions

- The dataset accurately represents real-world news and fake news.
- Each news article is independent and does not rely on social network propagation.
- Preprocessing removes only irrelevant or redundant data without affecting essential information.
- Classifiers are trained on sufficient labeled data for effective learning.

III. EFFICIENT COMMUNICATION

Efficient communication plays a vital role in the proposed fake news detection system, as it ensures fast, reliable, and secure data flow between different system components. The system is designed to minimize processing delays while handling large volumes of news data in real time.

The communication process begins with the transfer of news articles from the data source to the preprocessing module. Lightweight data formats and optimized text pipelines are used to reduce transmission overhead. During preprocessing and feature extraction, only relevant textual information is passed forward, which improves computational efficiency and reduces memory usage.

Machine learning models are trained offline to avoid communication overhead during real-time prediction. Once trained, the models efficiently receive feature vectors and return classification results with minimal latency. This design enables quick decision-making and supports real-time or near real-time fake news detection.

Additionally, the system supports secure communication protocols to prevent unauthorized access and data tampering. Efficient communication combined with security mechanisms ensures reliable system performance, scalability, and accurate classification results even under high data traffic conditions.

IV. SECURITY

Security is a critical aspect of the proposed fake news detection system, as it handles large volumes of digital news data and machine learning models. Ensuring data confidentiality, integrity, and availability is essential to maintain system reliability and trustworthiness.

To protect sensitive data, the system employs secure storage mechanisms and access control policies that restrict unauthorized users from modifying or accessing datasets and trained models. Encryption techniques are applied during data storage and transmission to prevent data leakage and interception.

The system also incorporates authentication and authorization mechanisms to ensure that only verified users and services can interact with the classification modules. This helps prevent malicious attacks such as data poisoning, where attackers attempt to manipulate training data to affect model performance.

Additionally, regular monitoring and logging mechanisms are used to detect abnormal activities and ensure system integrity. By combining secure communication protocols with controlled access and monitoring, the proposed system provides a robust and secure environment for automated fake news detection.

V. MODELING AND ANALYSIS

This section describes the modeling approach and analytical process used for fake news detection. The system is modeled as a supervised text classification problem, where news articles are represented using textual features and classified as real or fake. After data collection, the textual content is preprocessed through tokenization, stop-word removal, and normalization to improve data quality. Feature extraction techniques such as Bag of Words (BoW) and Term Frequency–Inverse Document Frequency (TF-IDF) are applied to convert text into numerical vectors suitable for machine learning algorithms.

Multiple supervised learning models, including Naïve Bayes, Logistic Regression, and Support Vector Machines (SVM), are trained on the extracted features. Model performance is analyzed using standard evaluation metrics such as accuracy, precision, recall, and F1-score. Comparative analysis is performed to identify the most effective model based on classification performance and computational efficiency. The results indicate that text-based machine learning models can effectively capture linguistic patterns associated with fake news. This modeling and analysis validate the suitability of the proposed approach for automated fake news detection.

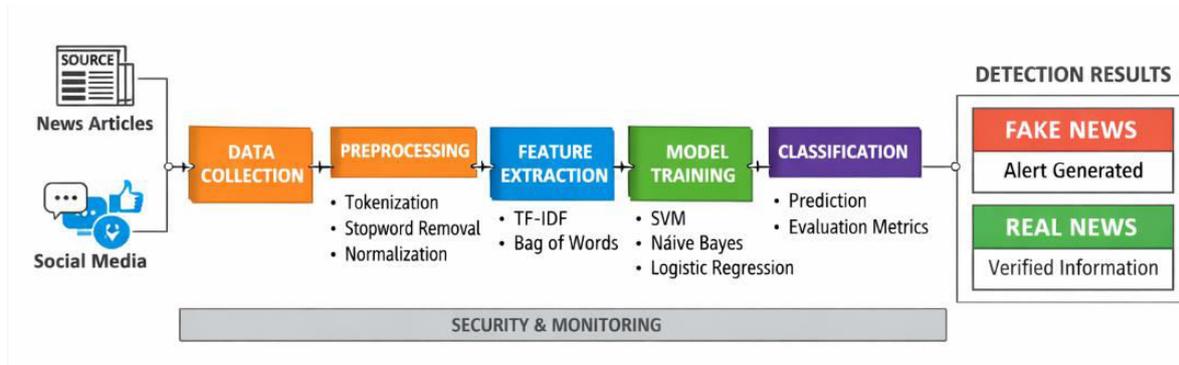


Fig. System Block Diagram

VI. CONCLUSION

In today's digital era, the rapid spread of false information poses serious challenges to public opinion and trust. This study presented a machine learning-based system for detecting fake news, utilizing text preprocessing, feature extraction, and supervised classification models such as Naïve Bayes, Logistic Regression, and Support Vector Machines. The results demonstrate that automated detection techniques can achieve reliable accuracy in identifying misinformation. The system also emphasizes secure data handling and clear presentation of results, ensuring trustworthiness and usability. Future work may include incorporating multimodal data, improving real-time detection, and enhancing model robustness against evolving forms of misinformation. Overall, the proposed framework provides a practical and effective approach to mitigate fake news and support informed decision-making.

REFERENCES

1. Wang, W. Y. (2017). A benchmark dataset for fake news detection. *Proceedings of the ACL*, 422–426.
2. Rashkin, H., et al. (2017). Linguistic analysis of misinformation and fact-checking content. *EMNLP Proceedings*, 2931–2937.
3. Zhou, X., & Zafarani, R. (2020). Fake news detection: Methods and challenges. *ACM Computing Surveys*, 53(5), 1–40.
4. Shu, K., et al. (2019). FakeNewsNet: A data repository for fake news research. *Big Data Journal*, 7(1), 41–56.
5. Kaliyar, R. K., et al. (2021). Detection of fake news using deep learning techniques. *Multimedia Tools and Applications*, 80(8), 11765–11788.



INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 9940 572 462  6381 907 438  ijirset@gmail.com



www.ijirset.com

Scan to save the contact details