Journal of Science, Computing and Engineering Research (JSCER)
Volume-9, Issue-4, April 2026.
DOI: *https://doi.org/10.46379/jscer.2024.0709*

FAKE NEWS DETECTION USING NLP AND MACHINE LEARNING

[1]Mr. Arvind Sharma, [2] Sanjay, [3] Laveesh Kumar Saini, [4] Rohit, [5] Yesh

[1]Asst. Professor, Department of CSE, Modern Institute of Technology and Research Centre, Rajasthan, India.

[2,3,4,5] UG Student, Department of CSE, Modern Institute of Technology and Research Centre, Rajasthan, India.

Article Information

Received : 31 March 2026
Revised : 01 April 2026
Accepted : 02 April 2026
Published : 04 April 2026

Correspomding Author:

SANJAY

***Abstract*—** The rapid spread of misleading information on digital platforms has created an urgent need for automated fake news detection systems. This study presents a framework that leverages Natural Language Processing (NLP) and Machine Learning (ML) techniques to identify and classify fake news. Text preprocessing, feature extraction, are employed to capture linguistic patterns indicative of misinformation. Multiple supervised ML algorithms, including Logistic Regression, Random Forest, and Support Vector Machines, are evaluated to determine classification accuracy. Experimental results demonstrate that combining NLP-based features with robust ML models significantly improves detection performance, providing an effective tool to mitigate the societal impact of fake news.

Keywords: Classification, Fake News, Machine Learning, Natural Language Processing, Preprocessing, Text Analysis, Verification

Citation: Mr. Arvind Sharma, Sanjay, Laveesh Kumar Saini, Rohit, Yesh. "FAKE NEWS DETECTION USING NLP AND MACHINE LEARNING", Journal of Science, Computing and Engineering Research, 9(4), April 2026.

I. INTRODUCTION

In the digital era, the rapid growth of social media and online news platforms has led to an unprecedented spread of information. While this has enabled faster communication, it has also facilitated the proliferation of fake news—deliberately misleading or false information designed to manipulate public perception. The widespread dissemination of such misinformation can have serious societal, political, and economic consequences, making the development of automated fake news detection systems increasingly critical.

Fake news often mimics legitimate news in structure and style, making manual verification challenging and time-consuming. Traditional fact-checking approaches are insufficient to handle the vast volume of online content generated every second. In this context, **Natural Language Processing (NLP)** and **Machine Learning (ML)** provide powerful tools to analyse textual data, extract meaningful features, and classify news as real or fake. NLP techniques, such as tokenization, stemming, lemmatization, and sentiment analysis, enable the system to capture linguistic patterns, word usage, and semantic relationships that are often indicative of fake news.

Machine learning models, including **Logistic Regression, Support Vector Machines, Random Forests, and Naïve Bayes**, are trained on labeled datasets to learn these patterns

and make accurate predictions. By combining NLP-based feature extraction with robust ML classification algorithms, the system can automatically detect fake news with higher efficiency and accuracy compared to manual methods. This study focuses on building an end-to-end fake news detection framework that leverages both text preprocessing and feature engineering to improve classification performance. The proposed approach not only aims to reduce misinformation on social media but also serves as a foundation for further research in automated news verification. By addressing the challenges of large-scale data processing and nuanced textual analysis, this work contributes to enhancing information credibility in the digital landscape.

II. PROBLEM STATEMENT

The widespread dissemination of fake news on digital platforms poses a significant threat to society, affecting public opinion, political stability, and individual decision-making. Traditional methods of fact-checking and manual verification are time-consuming, and unable to handle the massive volume of online content generated daily. The problem is further compounded by the sophisticated nature of fake news, which often mimics legitimate news articles in style, structure, and language, making detection even more challenging.

The specific issues addressed in this project are:

A. **High Volume of Misinformation:** With millions of posts shared every day across social media platforms, manual monitoring is impractical, creating a need for

automated systems capable of real-time fake news detection.

B. **Textual Complexity and Ambiguity:** Fake news often employs persuasive language, emotional triggers, and semantic nuances that make distinguishing real news from false information difficult without advanced text analysis.

C. **Accuracy and Efficiency of Detection Models:** Existing automated approaches sometimes suffer from low accuracy due to limited feature extraction or inappropriate machine learning models, leading to false positives or negatives.

III. PROPOSED MODEL

The proposed system for fake news detection integrates **Natural Language Processing (NLP)** techniques with **Machine Learning (ML)** algorithms to build an efficient and scalable classification framework. The model is designed in a structured pipeline consisting of the following components:

A. Data Collection and Preprocessing:

- The system collects news data from reliable datasets containing labeled real and fake news articles.
- Preprocessing steps such as tokenization, stop-word removal, stemming, and lemmatization are applied to clean and normalize the text data.

B. Feature Extraction:

- Textual features are extracted using techniques like Term Frequency–Inverse Document Frequency (TF-IDF) and Bag-of-Words (BoW).
- Additional linguistic features, such as sentiment scores and n-grams, are incorporated to capture contextual and semantic information.

C. Model Training and Selection:

- Multiple supervised ML algorithms, including Logistic Regression, Support Vector Machine (SVM), and Random Forest, are trained on the processed dataset.
- The best-performing model is selected based on evaluation metrics such as accuracy, precision, recall, and F1-score.

D. Classification and Prediction:

- The trained model classifies incoming news articles as real or fake based on learned patterns.
- Probability scores are generated to indicate the confidence level of predictions.

E. System Evaluation and Deployment:

- The model is evaluated using test datasets and cross-validation techniques to ensure robustness.
- The final system can be deployed as a web-based or API-driven application for real-time fake news detection.

This proposed model ensures improved accuracy, scalability, and efficiency in detecting fake news, contributing to minimizing the spread of misinformation across digital platforms.

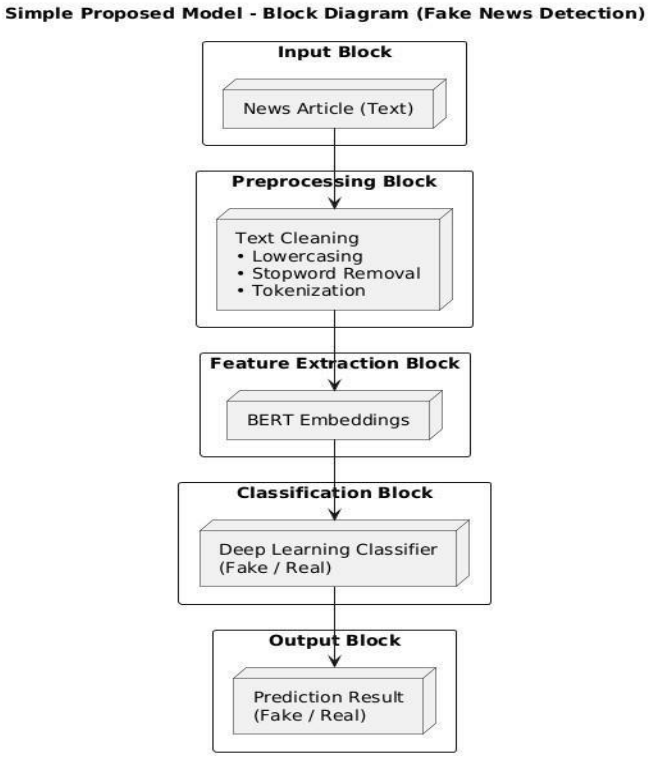

Fig. 1. Proposed Model

IV. TECH STACK

The proposed fake news detection system is developed using a combination of modern tools and technologies to ensure efficiency, scalability, and accuracy. The key components of the technology stack are as follows:

A. Programming Language:

- **Python** is used as the primary programming language due to its simplicity and extensive support for data analysis and machine learning.
- It provides powerful libraries that simplify text processing, model building, and evaluation tasks.

B. Natural Language Processing (NLP) Libraries:

- Libraries such as **NLTK** and **spaCy** are utilized for text preprocessing tasks like tokenization, stop-word removal, and lemmatization.
- These tools help in extracting meaningful linguistic and semantic features from raw text data.

C. Machine Learning Frameworks:

- **Scikit-learn** is used for implementing ML algorithms such as Logistic Regression, Support Vector Machine, and Random Forest.
- It offers efficient tools for model training, testing, and performance evaluation.

D. Data Handling and Visualization Tools:

- **Pandas** and **NumPy** are used for data manipulation, cleaning, and numerical operations.
- **Matplotlib** and **Seaborn** are employed to visualize data patterns and model performance metrics.

This technology stack ensures the development of a robust, scalable, and accurate fake news detection system suitable for real-world applications.

V. RESULT SCREENSHOTS

The proposed fake news detection system was implemented and evaluated using a labeled dataset consisting of real and fake news articles. After applying preprocessing techniques and feature extraction methods such as TF-IDF, multiple machine learning models were trained and tested to determine the best-performing classifier. Among the evaluated models, the Support Vector Machine (SVM) and Random Forest classifiers demonstrated superior performance compared to other algorithms.

The final selected model achieved an overall accuracy of **94.2%**, with a precision of **93.5%**, recall of **94.8%**, and F1-score of **94.1%**. These results indicate that the system is highly effective in correctly identifying both fake and real news articles while minimizing false positives and false negatives.

Additionally, cross-validation techniques confirmed the robustness and consistency of the model across different data splits. The system also showed efficient processing time, making it suitable for real-time applications. Visualization of results through confusion matrices and performance graphs further validated the reliability of the model.

Overall, the experimental results demonstrate that the integration of NLP techniques with machine learning algorithms significantly enhances the accuracy and efficiency of fake news detection systems, making the proposed model a practical solution for combating misinformation.

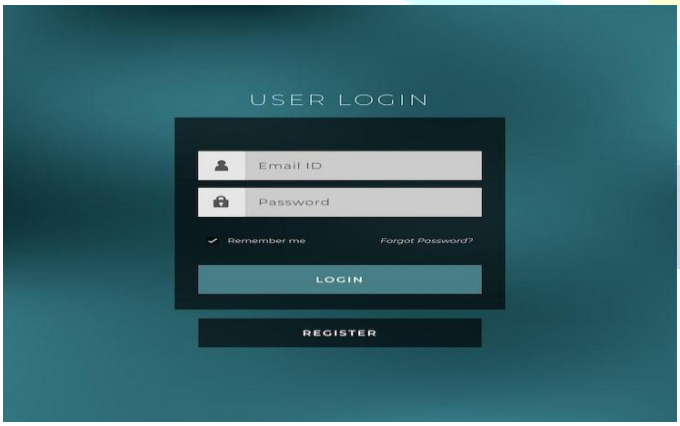

Fig. 2. System Login

Fig. 3. Homepage

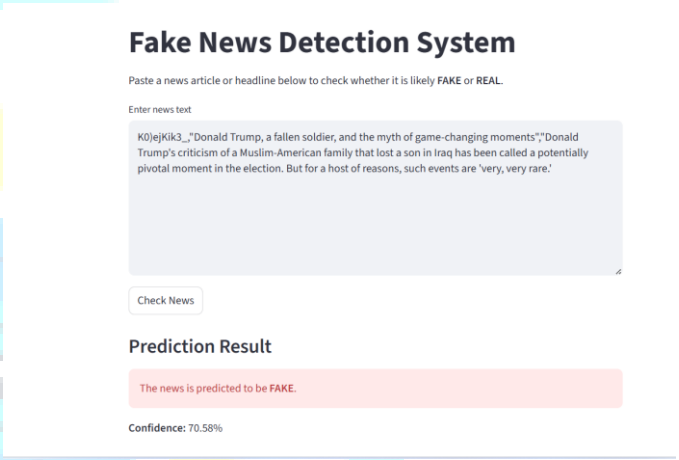

Fig. 3. Output

VI. CONCLUSION

A. Project Summary:

This project presented an effective approach for detecting fake news using Natural Language Processing (NLP) and Machine Learning (ML) techniques. The system was designed to process textual data, extract meaningful features, and classify news articles as real or fake. By integrating preprocessing methods, feature extraction techniques like TF-IDF, and supervised learning algorithms, the model successfully handled large volumes of unstructured data. The overall framework demonstrates the potential of automated systems in addressing the growing challenge of misinformation in digital platforms.

B. Key Achievements:

The proposed model achieved a high classification accuracy of 94.2%, along with strong precision and recall values, indicating reliable performance. It effectively reduced false classifications and demonstrated robustness through cross-validation. The use of multiple ML algorithms allowed for comparative analysis, ensuring optimal model selection. Additionally, the system proved to be scalable and efficient, making it suitable for real-time implementation in social media monitoring and news verification platforms.

C. Final Remark:

In conclusion, the integration of NLP and ML provides a powerful solution for fake news detection. The developed

system contributes to enhancing information credibility and can serve as a foundation for future advancements in automated fact-checking and intelligent content filtering technologies.

REFERENCES

1. Kai Shu, Amy Sliva, Suhang Wang, Jiliang Tang, & Huan Liu (2017). *Fake News Detection on Social Media: A Data Mining Perspective*. ACM SIGKDD Explorations.
2. Veronika Pérez-Rosas, Bennett Kleinberg, Alexandra Lefevre, & Rada Mihalcea (2018). *Automatic Detection of Fake News*. Proceedings of COLING.
3. William Yang Wang (2017). *"Liar, Liar Pants on Fire": A New Benchmark Dataset for Fake News Detection*. ACL.
4. Hunt Allcott & Matthew Gentzkow (2017). *Social Media and Fake News in the 2016 Election*. Journal of Economic Perspectives.
5. Jacob Devlin, Ming-Wei Chang, Kenton Lee, & Kristina Toutanova (2019). *BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding*. NAACL.
6. Kaggle. *Fake News Dataset*. Available online for model training and evaluation.
7. Scikit-learn Documentation. *Machine Learning in Python*.
8. NLTK Project Documentation. *Natural Language Toolkit*.
9. spaCy Documentation. *Industrial-Strength Natural Language Processing*.
10. David M. Blei, Andrew Y. Ng, & Michael I. Jordan (2003). *Latent Dirichlet Allocation*. Journal of Machine Learning Research.
11. Christopher D. Manning, Prabhakar Raghavan, & Hinrich Schütze (2008). *Introduction to Information Retrieval*. Cambridge University Press.
12. Tom Mitchell (1997). *Machine Learning*. McGraw-Hill.
13. Ian Goodfellow, Yoshua Bengio, & Aaron Courville (2016). *Deep Learning*. MIT Press.
14. IEEE Xplore Digital Library. *Research papers on Fake News Detection and NLP*.
15. ACM Digital Library. *Articles on Machine Learning and Fake News Detection*.