International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 13 Issue: 01 | Jan 2026 www.irjet.net p-ISSN: 2395-0072

Fake News Detection

Dr.C.P.Divate[1], Ms.N.R.Bhokre[2], Saksham S. Shrivastav[3], Vikram R. Sargar[4], Adarsh K. Yevale [5], Priyanka V. Chougule [6], Onkar M. Shinde [7]

[1]*Dean, Dept of Computer Engineering, Shri Ambabai Talim Sanstha's Sanjay Bhokare Group of Institute Miraj(poly), Maharashtra, India*

[2]*Lecturer, Dept of Computer Engineering, Shri Ambabai Talim Sanstha's Sanjay Bhokare Group of Institute Miraj(poly), Maharashtra, India*

[3,4,5,6,7]*Student Dept of Computer Engineering, Shri Ambabai Talim Sanstha's Sanjay Bhokare Group of Institute Miraj(poly), Maharashtra, India*

Abstract - *The rapid growth of social media and online news platforms has led to an unprecedented spread of misinformation and **fake news**, which can influence public opinion, create social unrest, and undermine trust in digital media. Manual verification of online content is time-consuming and impractical at a large scale. This paper presents a fake news detection system based on Natural Language Processing (NLP) and rule-based analysis techniques, without the use of machine learning models.*

Key Words: **fake news detection; natural language processing; rule-based system; text analysis; misinformation.**

1.INTRODUCTION

The consumption of news has shifted significantly from traditional print and broadcast media to digital platforms, including online news portals and social networking sites. This transformation has greatly enhanced the accessibility, immediacy, and reach of information, allowing users to receive real-time updates from across the globe. However, this same ease of distribution has also facilitated the rapid spread of fake, misleading, or deliberately manipulated news. Unlike traditional media, online platforms often lack rigorous editorial oversight, enabling unverified content to circulate widely within minutes.

Numerous studies have shown that false information tends to spread faster and reach a wider audience than verified news, as it often appeals to emotions, sensationalism, or confirmation bias.

This issue becomes particularly critical during sensitive events such as elections, public health emergencies, pandemics, and large-scale social movements, where misinformation can influence public opinion, create panic, undermine trust in institutions, and even lead to real-world consequences. As a result, the detection and prevention of fake news have become essential challenges in the digital age, highlighting the need for automated, reliable, and scalable solutions..

2. PROBLEM STATEMENT

The rapid expansion of digital news platforms and social media has significantly transformed the way information is produced, shared, and consumed. While this digital shift has made news easily accessible and enabled real-time information dissemination, it has also resulted in the widespread circulation of fake, misleading, and deliberately manipulated news content. Such misinformation can influence public opinion, distort facts, create social unrest, and generate panic among the public, especially during critical events such as elections, natural disasters, and public health emergencies.

Due to the massive volume of content generated every minute on online platforms, manual verification of news authenticity has become extremely time-consuming, labor-intensive, and impractical. Human fact-checkers cannot efficiently keep up with the speed and scale at which false information spreads across digital platforms. As a result, fake news often reaches a large audience before it can be identified and corrected.

Many existing fake news detection systems rely primarily on keyword-based filtering or shallow text analysis techniques. These traditional approaches focus on surface-level patterns rather than understanding the contextual and semantic meaning of news articles. Consequently, they often fail to detect sophisticated fake news that uses subtle language, emotional manipulation, or partially true information, leading to inaccurate or unreliable predictions.

Furthermore, most existing systems simply classify news as either real or fake without providing a confidence score or explanation for the decision. This lack of transparency makes it difficult for users to assess the credibility of the output or understand how reliable the classification is. Therefore, there is a critical need for an intelligent, accurate, and user-friendly fake news detection system that leverages advanced natural language processing and machine learning techniques to evaluate news credibility effectively while also providing meaningful confidence levels to users.

3. PROPOSED SOLUTION

The proposed solution involves the development of an automated fake news detection system that utilizes machine learning techniques to accurately classify news articles as either real or fake. The primary objective of the system is to reduce human involvement while achieving high accuracy, scalability, and adaptability to the continuously evolving patterns of misinformation present in digital media. By analyzing textual content along with linguistic and statistical features of news articles, the system is capable of distinguishing genuine information from misleading or fabricated content.

The system begins with the collection of data from reliable and publicly available datasets that contain labeled news articles categorized as real or fake. To ensure data quality and consistency, the collected data undergoes a preprocessing phase. This phase includes text normalization, removal of stop words, elimination of punctuation and special characters, tokenization, and stemming or lemmatization. These preprocessing steps help reduce noise in the data and enhance the effectiveness of the machine learning models.

After preprocessing, relevant features are extracted from the textual data using techniques such as Bag of Words (BoW), Term Frequency–Inverse Document Frequency (TF-IDF), and word embeddings. Feature extraction transforms unstructured text into numerical representations that can be efficiently processed by machine learning algorithms.

The system employs multiple supervised machine learning classifiers, including Naïve Bayes, Logistic Regression, Support Vector Machines (SVM), and Random Forest. These algorithms are trained using the extracted features to identify patterns commonly associated with fake and real news. A comparative analysis of the classifiers is performed to determine the most accurate and reliable model.

Once trained, the selected model is evaluated using standard performance metrics such as accuracy, precision, recall, and F1-score to ensure reliability and effectiveness. The finalized model is then integrated into a user-friendly application that allows users to input a news article or URL and receive real-time authenticity predictions.

To maintain robustness and long-term effectiveness, the system is designed to be scalable and continuously updatable with new datasets. This enables the model to adapt to emerging misinformation trends and evolving language patterns. Overall, the proposed solution offers an efficient, cost-effective, and automated approach to addressing the growing challenge of fake news in digital media.

4.LITERATURE REVIEW

1.Wang (2017) highlighted that textual cues such as word choice, sentence complexity, and emotional tone can serve as indicators of misinformation, even without advanced learning models. This work demonstrated the importance of linguistic features in identifying deceptive content.

2.Shu et al. (2020) emphasized the role of source credibilityand contextual information in fake news detection. Their studies showed that unreliable sources tend to publish content with similar linguistic and stylistic patterns, making rule-based source verification an effective approach.

Other studies explored sentiment analysis techniques and found that fake news articles often exhibit extreme positive or negative sentiment compared to factual reporting. Rule-based systems combining sentiment scores, keyword lists, and writing-style heuristics were shown to provide transparent and explainable detection mechanisms.

Survey papers conclude that while rule-based systems may lack adaptability, they offer advantages such as interpretability, low computational cost, and ease of implementation, making them suitable for educational, prototype, and low-resource environments.

Research Objectives:

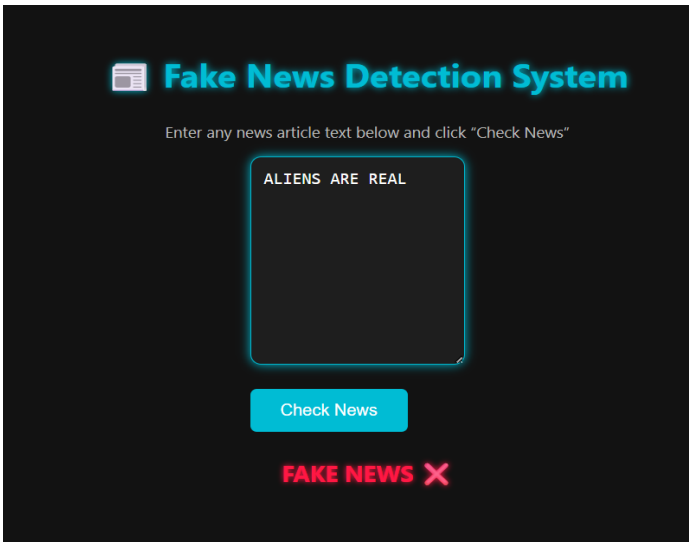
1. To implement Logistic Regression for efficient and accurate classification of news content.
2. To design and develop a machine learning–based model for detecting fake and real news automatically.
3. To evaluate the performance of the proposed model using standard accuracy and classification metrics.

4. To use Sentence-BERT (SBERT) for extracting meaningful semantic features from news text.

5.METHODOLOGY

The proposed Fake News Detection System follows a structured NLP-based pipeline without using machine learning models.

Initially, news articles are collected from online sources or user input. The text data undergoes preprocessing steps including conversion to lowercase, removal of punctuation, stop-word elimination, and tokenization to reduce noise and standardize input.



After preprocessing, multiple rule-based analyses are applied:

- **Keyword Analysis:** Detection of commonly used fake-news terms and sensational phrases
- **Sentiment Analysis:** Identification of extreme emotional polarity using lexicon-based methods
- **Writing Style Analysis:** Examination of headline length, punctuation usage, capitalization, and repetition
- **Source Credibility Check:** Verification against predefined lists of trusted and untrusted domains

Each analysis contributes to a cumulative credibility score. Based on predefined thresholds, the system classifies the news article as *real* or *potentially fake*. A simple user interface can be developed using Python-based tools to allow users to input news text and receive instant results.

6.Benefits of an NLP-Based Fake News Detection System

[1] A. User-Centric Benefits

An NLP-based fake news detection system offers significant advantages to end users by empowering them to make informed decisions while consuming digital content. One of the key benefits is accurate information verification, which allows users to quickly assess whether a news article is potentially fake or reliable. This improves user awareness and encourages critical thinking when interacting with online information.

The system also helps reduce exposure to misinformation by identifying suspicious content at an early stage. By discouraging users from sharing misleading or false news, it plays an important role in limiting the spread of misinformation across social media platforms.

Additionally, the system is designed with ease of use in mind. A simple, text-based interface ensures that even non-technical users can easily input news content and receive clear, understandable results, making the system widely accessible.

[2] B. Societal and Media Benefits

From a societal perspective, the system contributes to improving media credibility by supporting journalists, editors, and fact-checkers in preliminary content screening. Automated analysis helps identify potentially misleading articles before they are published or widely circulated.

The ability to perform real-time or near real-time analysis enables a faster response to misinformation, which is critical in preventing fake news from going viral. This is particularly valuable during sensitive events such as elections, public health crises, and social movements.

Furthermore, the system supports digital literacy by encouraging responsible content consumption and ethical sharing practices. By making users more aware of misinformation patterns, it promotes a healthier and more trustworthy digital information ecosystem.

[3] C. Technical Benefits

From a technical standpoint, the NLP-based rule-driven approach offers several advantages. The system operates with low computational cost, as rule-based NLP techniques require minimal processing power and can run efficiently on standard hardware without the need for expensive GPUs or large training datasets.

The scalable design allows the system to be easily integrated into websites, browser extensions, news portals, or social media platforms. This flexibility makes it suitable for deployment across various digital environments.

Finally, the transparent architecture of the system ensures that decisions are explainable and interpretable. Unlike black-box machine learning models, the rule-based approach allows users and developers to understand how conclusions are reached, increasing trust, reliability, and ease of maintenance.

7.Conclusion

The Fake News Detection System using Natural Language Processing presents a practical and efficient approach to identifying misinformation in online news content without relying on machine learning techniques. By employing rule-based methods, the system analyzes key linguistic features such as word usage patterns, sentiment polarity, writing style, and source credibility to determine the authenticity of news articles. This rule-based framework ensures transparency in decision-making, as each classification result can be traced back to clearly defined rules and linguistic indicators. Additionally, the system is computationally lightweight, making it suitable for real-time applications and deployment on platforms with limited computational resources.

The proposed system serves as a valuable tool for users, journalists, researchers, and online content platforms by helping to minimize the spread of fake news and enhance trust in digital information ecosystems. It enables users to make informed decisions by providing clear and understandable outputs, thereby promoting media literacy and responsible content consumption. The system's reliance on natural language processing techniques allows it to detect misleading patterns that may not be immediately apparent to human readers.

Despite its effectiveness, there is scope for further improvement. Future enhancements may include the incorporation of multilingual support to address misinformation across diverse linguistic communities. The system can also be extended by integrating hybrid statistical or lightweight learning-based techniques to improve adaptability while maintaining transparency. Additionally, deeper semantic analysis and contextual understanding can be introduced to better capture subtle misinformation, sarcasm, and evolving language patterns. These enhancements would further improve detection accuracy, robustness, and scalability, making the system more effective in combating the growing challenge of fake news in the digital era.

8.ACKNOWLEDGEMENT

We would like to extend our sincere gratitude to all those who contributed to the completion of this project on **Fake News Detection**. Special thanks to our mentors and advisors for their valuable guidance, continuous support, and encouragement throughout the development of this project.

Additionally, we sincerely appreciate the assistance of our colleagues and peers who provided helpful feedback, suggestions, and motivation during the course of this work. Their insights and cooperation greatly contributed to the successful completion of this project.

9.References

1] **Wang, W. Y.** "Liar, Liar Pants on Fire: A New Benchmark Dataset for Fake News Detection," *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (ACL)*, Vancouver, Canada, 2017.
This paper introduces the LIAR dataset, a widely used benchmark for fake news detection research. It provides labeled short political statements along with metadata, enabling the evaluation of various natural language processing and classification approaches.

2] **Shu, K., Sliva, A., Wang, S., Tang, J., and Liu, H.** "Fake News Detection on Social Media: A Data Mining Perspective," *ACM SIGKDD Explorations Newsletter*, vol. 19, no. 1, pp. 22–36,2017.
This work presents a comprehensive survey of fake news detection methods from a data mining perspective, covering content-based, social context-based, and hybrid approaches, and highlights key challenges in combating misinformation.

3] **Shu, K., Mahudeswaran, D., Wang, S., Lee, D., and Liu, H.** "FakeNewsNet: A Data Repository with News Content,

Social Context, and Dynamic Information for Studying Fake News on Social Media," *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI)*, 2020. This paper introduces FakeNewsNet, a large-scale dataset that combines news content with social engagement data, enabling advanced research on fake news propagation and detection.

4] **Allcott, H., and Gentzkow, M.** "Social Media and Fake News in the 2016 Election," *Journal of Economic Perspectives* This study analyzes the role of social media in spreading fake news during the 2016 U.S. presidential election and examines its impact on public opinion and democratic processes.

5] **Vosoughi, S., Roy, D., and Aral, S.** "The Spread of True and False News Online," *Science*, vol. 359, no. 6380, pp. 1146–1151,2018.
This influential research demonstrates that false news spreads faster and more widely than true news on social media platforms, primarily due to human behavior rather than automated bots.

6] **Rashkin, H., Choi, E., Jang, J. Y., Volkova, S., and Choi, Y.** "Truth of Varying Shades: Analyzing Language in Fake News and Political Fact-Checking," *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*,2017.
This paper explores linguistic differences between fake, real, and fact-checked news, providing insights into language patterns useful for rule-based and NLP-driven fake news detection systems.

7] **Reuters Institute.** *Reuters Handbook of Journalism*, Reuters Institute for the Study of Journalism, University of Oxford,2021.
This handbook outlines professional journalistic standards, ethics, and verification practices, serving as a reference for evaluating news credibility and source reliability.

8] **NLTK Project.** "Natural Language Toolkit (NLTK) Documentation."
NLTK is a widely used Python library for natural language processing tasks such as tokenization, stemming, sentiment analysis, and linguistic feature extraction, supporting rule-based text analysis systems.

9] **Python Software Foundation.** "Python Programming Language Documentation." Python provides a flexible and powerful programming environment for implementing fake news detection systems, offering extensive libraries for text processing, data analysis, and application development.

10]**Kaggle.** "Fake News Dataset. "Kaggle hosts publicly available fake news datasets that are commonly used for training, testing, and benchmarking fake news detection models and rule-based systems.