As per UGC guidelines an electronic bar code is provided to secure your paper

International Journal for Modern Trends in Science and Technology
Volume 12, Issue 04, pages 444-448.
ISSN: 2455-3778 online
Available online at: http://www.ijmtst.com/vol12issue04.html
DOI: https://doi.org/10.5281/zenodo.19470641

Fake News Detection on Social Media Using NLP

A. Naga Teja Sree, B.Deepthi Nikhila, D.Aishwarya, B.Hasini, B.Sree Lalitha Devi, K.Swathi, M.Mounika

Department of Computer Science and Engineering, Gouthami Institute of Technology and Management for Women, Andhra Pradesh, India.

To Cite this Article
A. Naga Teja Sree, B.Deepthi Nikhila, D.Aishwarya, B.Hasini, B.Sree Lalitha Devi, K.Swathi & M.Mounika (2026). Fake News Detection on Social Media Using NLP. International Journal for Modern Trends in Science and Technology, 12(04), 440-443. https://doi.org/10.5281/zenodo.19470641

Article Info
Received: 10 March 2026; Revised: 02 April 2026; Accepted: 05 April 2026.

KEYWORDS

Fake News Detection, Natural Language Processing (NLP)

ABSTRACT

Media sites creates serious problems for the accuracy of information and public confidence. Differentiating between real and fake information has grown more difficult as billions of people actively consume and share content online. The use of Natural Language Processing (NLP) approaches for automated fake news identification is investigated in this research. To categorize news as authentic or fraudulent, machine learning and deep learning models assess a variety of linguistic, semantic, and contextual factors. We assess several NLP-based models, such as Transformer-based architectures like BERT, Bidirectional LSTM, and TF-IDF using Logistic Regression. According to experimental findings, deep learning techniques—in particular, BERT-based models—achieve better accuracy and generalization across a variety of datasets. The paper explores future paths for enhancing explainability and cross-domain adaptability while highlighting the potential of NLP in reducing misinformation.

INTRODUCTION

way individuals consume information has been completely transformed by the quick expansion of social media sites like Facebook, Reddit, and Twitter. However, the dissemination of fake news—erroneous or misleading material portrayed as fact—has also been made easier by this accessibility [1]. Fake news has the power to sway public opinion, affect elections, provoke violence, and harm people's reputations [2]. The enormous volume of data created every second makes traditional manual verification methods A branch of artificial intelligence (AI) called natural language processing (NLP) offers computational methods for processing and comprehending human language [4]. NLP models are able to recognize minor indicators of dishonesty or false information by utilizing linguistic patterns, contextual signals, and semantic linkages [5]. NLP's capacity to identify fake news with high accuracy has been further improved by recent developments in machine learning (ML) and deep learning (DL) [6]. This paper emphases on designing an NLP-based framework for fake news detection on social media. The method

mixes feature extraction practices such as word embeddings and transformer encodings with supervised learning algorithms to classify news items. The proposed system is evaluated on benchmark datasets to measure its effectiveness and reliability.

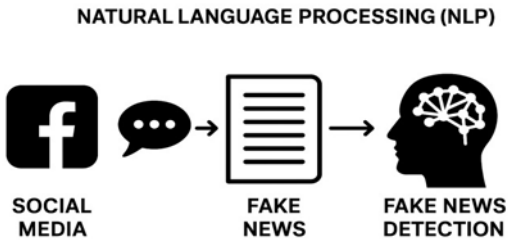

Figure1: Natural Language Processing (NLP)

The above figure 1 demonstrates the general process of fake news detection on social media using Natural Language Processing (NLP). It begins with data shared across social media platforms, where users post and circulate news content. This content, which may contain both honest and ambiguous information, is represented as textual data in the form of news articles, uprights, or comments. The next step includes NLPbased processing, where language and contextual features of the text are analysed using machine learning and deep learning techniques [7]. Lastly, the handled information is passed through a fake news detection model, which identifies and classifies the content as either real or fake, thereby helping to battle misinformation and ensure more reliable information dissemination online.

RELATED WORK

Due to the exponential growth in social media usage and the subsequent spread of false information, the identification of fake news has drawn more attention in recent years. Manual fact-checking and rule-based systems, which depended on human knowledge and predetermined linguistic patterns to confirm the veracity of news, were the main focus of early research efforts. Although these methods offered a certain level of accuracy, their scalability and adaptability were intrinsically constrained since human fact-checking was unable to keep up with the enormous volume and speed of social media information. In order to get over these limitations, scientists started investigating automated approaches that use machine learning (ML) and natural language processing (NLP) techniques to examine linguistic clues and textual patterns suggestive of misleading or deceptive information.

A.Bag-of-WordsandTF-IDF Approaches

Traditional text representation methods like Bag-of-Words (BoW) and Term Frequency–Inverse Document Frequency (TF-IDF) were the mainstay of the first generation of automated false news detection systems. These techniques ignore grammatical and contextual linkages in favor of representing documents as numerical feature vectors based on word occurrences. By examining linguistic patterns and rhetorical structures in text, Rubin et al. (2016) developed one of the first machine learning-based methods for identifying false news. This was further developed by Potthast et al. (2018) using a large dataset and traditional machine learning classifiers including Support Vector Machines (SVM), Naïve Bayes, and Logistic Regression. According to their findings, BoW and TF-IDF characteristics might offer a modest level of classification accuracy, usually between 70% and 80%. However, these models were constrained by their incapacity to capture contextual dependencies, polysemy (words with many meanings), and semantic nuances—all of which are essential for identifying the subtle linguistic manipulations frequently found in false news.

B. Deep Learning Models

Researchers started using neural network-based architectures that could more successfully capture semantic and contextual information as deep learning progressed [8]. In order to understand both spatial(local) and sequential (temporal) relationships in text data, Ruchansky et al. (2017) presented the CSI model, a hybrid deep learning framework that combines Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs). One of the earliest attempts to combine content-based and socialcontext elements was this model, which also included user behavior and temporal engagement patterns [9]. In a similar vein, Long Short-Term Memory (LSTM) networks—an extension of RNNs—performed better while managing textual long-range dependencies. In order to improve classification accuracy, studies like Wang (2017) and Singhal et al. (2019) used bidirectional LSTMs (BiLSTMs) to assess news items from both forward and backward

contextual perspectives. These models were able to replicate the complex grammatical irregularities, emotional tones, and writing patterns found in fake news articles. However, deep learning methods hampered their generalizability and interpretability across a variety of areas because they required extensive computational resources and big annotated datasets [10][11].

C.Transformer-Based Models

An important development in NLP-based fake news detection was the introduction of Transformer architectures. Vaswani et al. (2017) introduced transformers, which depend on self-attention mechanisms to enable models to capture global identifying minute contextual anomalies in news information because it can comprehend language bidirectionally, taking into account both left and right context [12].

BERT and its variations, such RoBERTa and XLNet, consistently beat conventional and RNN-based models across a variety of false news datasets, including FakeNewsNet, LIAR, and Kaggle false News Challenge, according to later research by Zhou and Zafarani (2020) and others. These transformer-based models successfully modeled syntactic and semantic dependencies and adjusted to linguistic variation, achieving state-of-the-art performance with accuracies surpassing 90%.

D.Research Gaps

While previous studies have achieved significant progress, several gaps remain:

• Domain Adaptation: Models trained on specific topics (e.g., politics) often fail on other domains (e.g., health misinformation).

• Explainability: Deep and transformer-based models often act as ―black boxes,‖ providing limited interpretability.

• Cross-Platform Generalization: Fake news patterns differ across platforms (Twitter vs. Facebook), reducing model transferability.

• Multimodality: Most studies focus solely on textual content, ignoring images, videos, or metadata that may provide additional cues. In order to improve generalization, interpretability, and performance consistency [13], this study compares transformer-based, deep learning, and conventional NLP models across several social media datasets.

METHODOLOGY

The following figure2 demonstrates the methodology framework for the research on Fake News Detection on social media Using NLP. It characterizes a step-bystep flow of how data is processed, analysed, and evaluated to detect fake news automatically.

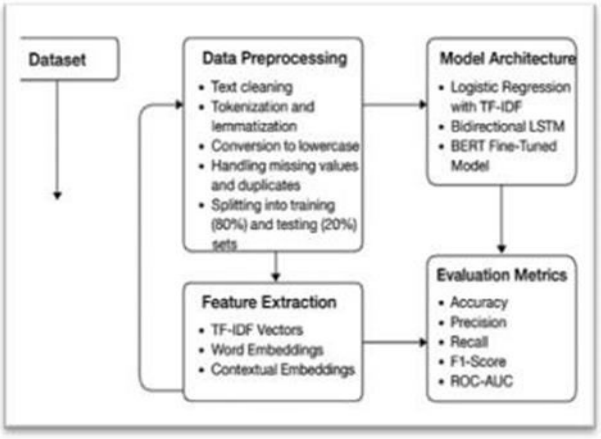

Figure2:Methodology of Fake news detection

The method starts with the Dataset, which comprises benchmark datasets that are accessible to the public, like FakeNewsNet, LIAR, and the Kaggle Fake News dataset [14]. Examples of both authentic and fraudulent news stories are identified in these databases. Cleaning and preparing the text for analysis is the next stage, known as data preprocessing [15]. This entails eliminating extraneous elements like punctuation and URLs, tokenizing and lemmatizing words, translating all text to lowercase, handling missing values, and dividing the data into training (80%) and testing (20%) groups [16].

After preprocessing, the data moves to the Feature Extraction stage, where textual material is transformed into numerical illustrations suitable for machine learning [17]. Three types of features are extracted: TF-IDF vectors for term importance, word embeddings (Word2Vec, GloVe) for semantic understanding, and contextual embeddings using BERT to capture deeper language context [18].

The various models assessed in this study are displayed in the following block, Model Architecture: a refined BERT model (a transformer-based deep contextual model), Bidirectional LSTM (a deep learning model for sequential text analysis), and Logistic Regression with TF-IDF (a conventional machine

learning baseline). These models pick up patterns that differentiate between authentic and fraudulent news. Lastly, the outputs are analysed using Evaluation Metrics such as Accuracy, Precision, Recall, F1-Score, and ROC-AUC, which deliver a quantitative assessment of model performance [19]. Composed, this methodology ensures a systematic and data-driven approach to fake news detection using NLP and modern AI techniques [20].

RESULTS AND DISCUSSION

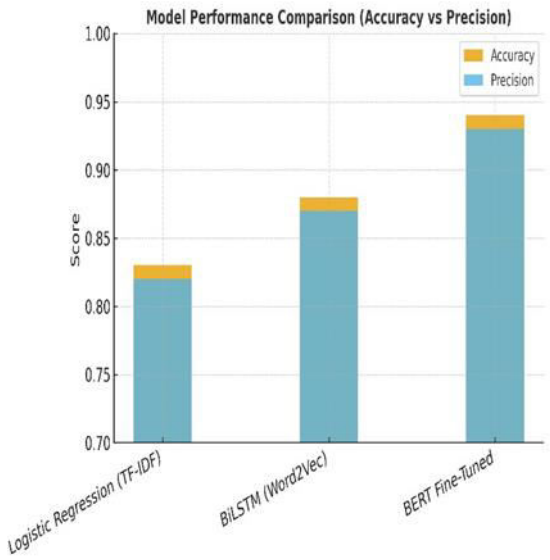

Figure 3: Model performance comparison

When compared to conventional models, the BERT model outperformed them in terms of contextual comprehension [21]. Although it needed more processing power, the BiLSTM model also did well [22]. Despite being straightforward and effective, the Logistic Regression baseline lacked semantic depth [23][24].
According to qualitative study, the BERT model was successful in identifying fake news, which is defined by sensational language, exaggerated sentiment, and a lack of reliable sources. Nevertheless, dealing with irony, humor, and domain-specific vocabulary continues to provide difficulties.

CONCLUSION

The effectiveness of NLP approaches in identifying false information on social media platforms is demonstrated by this study. Transformer-based designs, like BERT, demonstrated the best overall performance among the assessed models because of their profound contextual understanding. When compared to conventional machine learning techniques, the incorporation of syntactic and semantic data greatly improves detection accuracy.

Future research should concentrate on integrating multimodal data (text, photos, and metadata), enhancing model interpretability, and detecting crosslingual fake news. Online misinformation mitigation systems could also be strengthened by creating lightweight models for real-time social media monitoring.

Conflict of interest statement

Authors declare that they do not have any conflict of interest.

REFERENCES

[1] Thampi, S. M. (2019, December). User recognition using cognitive psychology based behavior modeling in online social networks. In International Symposium on Signal Processing and Intelligent Recognition Systems (pp. 130-149). Singapore: Springer Singapore.

[2] Li, X., Xiong, H., Li, X., Wu, X., Zhang, X., Liu, J., ... & Dou, D. (2022). Interpretable deep learning: Interpretation, interpretability, trustworthiness, and beyond. Knowledge and Information Systems, 64(12), 3197-3234.

[3] Kale, D. R., & Todmal, S. R. (2014). A survey on big data mining applications and different challenges. Int J Adv Res Comput Eng Technol, 3, 3835-3838.

[4] Choudhary, A., & Arora, A. (2024). Assessment of bidirectional transformer encoder model and attention based bidirectional LSTM language models for fake news detection. Journal of Retailing and Consumer Services, 76, 103545.

[5] Kale, M. D. R., & Todmal, M. S. R. (2015). A Result Paper on Investigation of Incremental Detection Problems in Distributed Data. International Journal of Advanced Research in Computer Engineering & Technology (IJARCET), 4(12).

[6] Sharma, D. K., & Garg, S. (2023). IFND: a benchmark dataset for fake news detection. Complex & intelligent systems, 9(3), 2843-2863.

[7] Kale, D. R., Dixit, C., Patil, N., & Kaulwar, B. (2025, October). Detection Systems for Robust IoT Security. In Integration of Artificial Intelligence in IoT: Opportunities and Challenges:Proceedings of 9th International Conference on Internet of Things].

[8] Kale, D. R., Nalvade, J., Randive, P. S., & Hirve, S. (2024). Artificial intelligence in sustainable agriculture: Enhancing efficiency and reducing and Connected Technologies (ICIoTCT 2024) (p. 123). Springer Nature.

[9] Yadav, A., Karnatak, V., Kale, D. R., & Chopra, M. Regression Based Intelligent Mechanism For Prediction Of Stock Values In Real-Time Invision.

[10] Kale, D., Jadhav, A., Wagh, M., Patil, S., Khatawkar, S., Patel, P., Maniyar, K., Mishra, E., & Patil, G. (2025). Quantum-Enhanced Big Data Analytics for Climate Change Predictions: A Scalable Solution

for Global Challenges. Journal of Mines, Metals and Fuels, 73(11), 3563–3575.

[11] Johnson, S. J., Murty, M. R., & Navakanth, I. (2024). A detailed review on word embedding techniques with emphasis on word2vec. Multimedia Tools and Applications, 83(13), 37979-38007.

[12] Kale, D. R., Jadhav, A. N., Salunkhe, S. J., Hirve, S., & Goswami, C. (2024, October). Sharding: a scalability solutions for blockchain networks. In 2024 IEEE International Conference on Blockchain and Distributed Systems Security (ICBDS) (pp. 1-8). IEEE.