IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 6, June 2026
Impact Factor: 8.2

Fake News Detection Using Natural Language Processing (NLP)

Gadade Vaishnavi Sanjay and Prof. Dr. Satish N. Gujar
Department of Computer Application
JSPM University, Pune
vaishnavigadade2003@gmail.com and sng.scos@jspmuni.edu.in

Abstract: *The rapid growth of social media and online news platforms has made information easily accessible, but it has also increased the spread of fake news. Fake news refers to false or misleading information presented as real news, which can influence public opinion, create confusion, and harm society. Detecting fake news manually is difficult because a huge amount of content is generated every day. Therefore, automated fake news detection systems are becoming important. This research paper focuses on detecting fake news using Natural Language Processing (NLP) techniques. NLP helps computers understand and analyze human language by examining the text content of news articles. In this study, different text processing methods such as tokenization, stop-word removal, stemming, and vectorization are used to prepare the data for analysis. Machine learning algorithms are then applied to classify news as real or fake based on linguistic patterns and writing styles. The proposed system aims to improve the accuracy and speed of fake news identification. By using NLP and machine learning together, the model can analyze large amounts of textual data efficiently and provide reliable predictions. This research can help reduce the spread of misinformation on digital platforms and support users in identifying trustworthy news sources. The study also highlights the importance of artificial intelligence in solving real-world communication and media-related problems. Based on language patterns and writing style, the model classifies news as real or fake. The proposed approach helps in reducing misinformation and improving the reliability of online news content. This study highlights the importance of NLP and artificial intelligence in modern news verification systems.*

Keywords: *Natural Language Processing*

I. INTRODUCTION

In the modern digital era, the internet has become one of the main sources of information for people around the world. Social media platforms, online news websites, blogs, and messaging applications allow users to access and share news instantly. Although this rapid flow of information has many advantages, it has also increased the spread of fake news. Fake news refers to false, misleading, or manipulated information that is presented as real news with the intention of influencing public opinion, creating confusion, or gaining financial and political benefits. The spread of fake news can negatively affect society by creating misunderstandings, fear, social conflict, and mistrust among people. The problem of fake news has grown significantly because social media platforms allow anyone to publish and share content without proper verification. News spreads very quickly online, and many users share information without checking whether it is true or false. During important events such as elections, natural disasters, health emergencies, and political campaigns, fake news can create serious consequences. Therefore, identifying and stopping the spread of fake news has become an important challenge in today's digital communication environment. Traditional methods of detecting fake news mainly depend on human fact-checkers and journalists. Although these methods can provide accurate verification, they require a large amount of time and effort. Since millions of news articles and social media posts are generated every day, manual verification alone is not enough. Because of this limitation, researchers and technology experts are developing automated fake news detection systems using Artificial Intelligence (AI), Machine Learning (ML), and Natural

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 6, June 2026
Impact Factor: 8.2

Language Processing (NLP). The main objective of this research is to develop an effective fake news detection system using NLP and machine learning techniques. The proposed system aims to analyze news articles automatically and classify them as real or fake with high accuracy and speed. Such systems can help reduce the spread of misinformation, improve the reliability of online information, and support users in making informed decisions. This research is important because fake news not only affects individuals but also impacts society, politics, economics, and public safety. By combining NLP and machine learning technologies, fake news detection systems can play a major role in maintaining trustworthy digital communication.

II. LITERATURE REVIEW

Fake news detection has become an important research area because of the rapid increase in online misinformation. Many researchers have worked on developing automated systems that can identify fake news using machine learning and Natural Language Processing (NLP) techniques. Different studies focus on analyzing the textual content of news articles, user behavior, and social media activities to improve detection accuracy. Early fake news detection methods mainly depended on manual fact-checking performed by journalists and experts.

Although these methods were reliable, they required significant time and human effort. As the volume of online content increased, researchers started using machine learning techniques to automate the detection process. Traditional machine learning algorithms such as Naive Bayes, Decision Tree, Logistic Regression, and Support Vector Machine (SVM) were widely used for classifying news articles as real or fake. These algorithms produced good results when trained on properly labeled datasets. Natural Language Processing plays a major role in fake news detection because it helps computers understand human language. Researchers have applied NLP techniques such as tokenization, stop-word removal, stemming, lemmatization, and text vectorization to preprocess news articles before classification. Feature extraction methods like Bag of Words (BoW) and Term Frequency–Inverse Document Frequency (TF-IDF) are commonly used to convert textual data into numerical format for machine learning models.Several studies have shown that linguistic features such as writing style, emotional tone, sentence structure, and word frequency can help distinguish fake news from genuine news. Fake news articles often contain sensational headlines, exaggerated language, and emotionally attractive words to gain user attention. Researchers have used sentiment analysis and semantic analysis to identify these patterns more effectively. In recent years, deep learning approaches have gained popularity in fake news detection. Models such as Recurrent Neural Networks (RNN), Long Short-Term Memory (LSTM), and Convolutional Neural Networks (CNN) can automatically learn complex language patterns from large datasets. These methods often provide higher accuracy compared to traditional machine learning techniques because they can capture contextual and sequential information in text data. Researchers have also explored transformer-based models such as BERT (Bidirectional Encoder Representations from Transformers) for fake news detection. These advanced NLP models understand the context and meaning of words more accurately, which improves classification performance. Many studies report that transformer-based models achieve better results in detecting misleading and manipulated content compared to earlier methods.

III. RESEARCH METHODOLOGY

Fake news detection has become an important research area because of the rapid increase in online misinformation. Many researchers have worked on developing automated systems that can identify fake news using machine learning and Natural Language Processing (NLP) techniques. Different studies focus on analyzing the textual content of news articles, user behavior, and social media activities to improve detection accuracy. Early fake news detection methods mainly depended on manual fact-checking performed by journalists and experts. Although these methods were reliable, they required significant time and human effort. As the volume of online content Natural Language Processing plays a major role in fake news detection because it helps computers understand human languageFeature extraction methods like Bag of Words (BoW) and Term Frequency–Inverse Document Frequency (TF-IDF) are commonly used to convert textual data into numerical format for machine learning models.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 6, June 2026
Impact Factor: 8.2

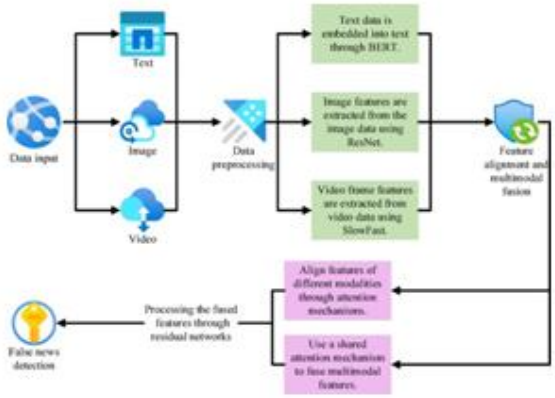
The table below summarizes key studies:

TABLE 1. SUMMARY OF KEY LITERATURE

Author	Year	Focus	Key Finding	Gap
Shu et al.	2017	Fake news detection using social media data	Social media behavior and news content can help identify fake news effectively.	Limited accuracy for rapidly changing news topics.
Rashkin et al.	2017	Linguistic analysis of fake news	Writing style and language patterns differ between fake and real news articles.	Does not fully analyze multimedia content such as images and videos.
Wang	2017	Machine learning for fake news classification	Machine learning models can classify fake news using textual features.	Small dataset reduces model generalization.

Several studies have shown that linguistic features such as writing style, emotional tone, sentence structure, and word frequency can help distinguish fake news from genuine news. Fake news articles often contain sensational headlines, exaggerated language, and emotionally attractive words to gain user attention. Researchers have used sentiment analysis and semantic analysis to identify these patterns more effectively. In recent years, deep learning approaches have gained popularity in fake news detection. Models such as Recurrent Neural Networks (RNN), Long Short-Term Memory (LSTM), and Convolutional Neural Networks (CNN) can automatically learn complex language patterns from large datasets.

These methods often provide higher accuracy compared to traditional machine learning techniques because they can capture contextual and sequential information in text data.



The first step in the research methodology involves collecting a suitable dataset for fake news detection. The dataset used in this research contains both fake and real news articles collected from publicly available sources such as Kaggle datasets, online repositories, and trusted news platforms. The collected dataset includes various fields such as news title, news content, category, and labels indicating whether the news is fake or real. The fake news articles contain misleading or manipulated information, whereas the real news articles are collected from reliable and verified news agencies. The dataset is selected carefully to maintain balance between fake and real news samples so that the machine learning model does not become biased toward one category. A balanced dataset improves the reliability and prediction capability of the system.

After collecting the dataset, the next stage is data preprocessing. Raw textual data obtained from online sources usually contains unnecessary information such as punctuation marks, special symbols, hyperlinks, numbers, HTML tags, stop words, and duplicate records. Such unwanted information can reduce the performance of the machine learning

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 6, June 2026
Impact Factor: 8.2

algorithms and increase computational complexity. Therefore, preprocessing is an essential step in fake news detection. During preprocessing, null values and incomplete records are removed from the dataset to maintain consistency. Duplicate news articles are also identified and deleted to avoid repeated learning by the model. The textual content is then cleaned by removing punctuation marks, special characters, URLs, extra spaces, and irrelevant symbols. This process helps simplify the text and improve the quality of the input data. All textual information is converted into lowercase letters to ensure uniformity across the dataset. For example, words such as "Government" and "government" are

Tier 1: Data Collection Layer

In this stage, news-related data is collected from multiple reliable and publicly available sources such as news websites, online repositories, Kaggle datasets, blogs, and social media platforms. The collected dataset contains both fake and real news articles that are required for training and testing the proposed machine learning model.

Tier 2: Data Processing and Integration Layer

The Data Processing and Integration Layer focuses on cleaning, organizing, and preparing the collected textual data for analysis. Raw news data collected from online platforms often contains noise such as punctuation marks, special symbols, hyperlinks, HTML tags, duplicate records, missing values, and irrelevant information.

Tier 3: Operational Performance Analysis

The Fake News Classification and Performance Analysis Layer is responsible for analyzing the processed data and identifying whether a news article is fake or genuine. In this stage, machine learning algorithms are trained using the transformed dataset to recognize textual patterns and linguistic characteristics associated with fake news.

Tier 4: Customer Insights and Business Intelligence

The Intelligent Insights and Decision Support Layer represents the final stage of the proposed framework. This layer focuses on generating meaningful insights and supporting users in identifying trustworthy information. Based on the prediction results produced by the machine learning model, the system provides clear classification outputs indicating whether the news article is fake or real.

V. JURISDICTION COMPARATIVE ANALYSIS

The rapid growth of digital communication platforms and social media has significantly increased the spread of fake news and misinformation across the world. Different countries have introduced various laws, regulations, and technological approaches to control the circulation of false information and protect public trust. The legal framework, level of technological adoption, and government policies related to fake news detection vary from one jurisdiction to another. Therefore, a comparative analysis of different jurisdictions is important to understand how countries are addressing the challenge of fake news and how Artificial Intelligence (AI) and Natural Language Processing (NLP) technologies can support these efforts. This section presents a comparative analysis of fake news regulations, digital governance policies, and technological approaches adopted by different countries such as India, the United States, the European Union, and Singapore. The analysis helps identify similarities, differences, strengths, and limitations in handling fake news and misinformation. India has experienced a significant rise in fake news circulation due to the widespread use of social media platforms such as WhatsApp, Facebook, YouTube, and X (formerly Twitter). Fake news in India often spreads through forwarded messages, manipulated videos, political propaganda, and misleading online articles

Such misinformation has affected public opinion, elections, religious harmony, and public safety. To address this issue, the Government of India has introduced several legal and regulatory provisions under the Information Technology Act, 2000 and related cyber laws. Various sections of the Indian Penal Code (IPC) are also used to take action against

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 6, June 2026
Impact Factor: 8.2

individuals spreading harmful misinformation. The Ministry of Electronics and Information Technology (MeitY) and fact-checking organizations play an important role in monitoring online content and identifying false information. Social media platforms operating in India are required to comply with government regulations related to harmful digital content.

TABLE 2. COMPARATIVE JURISDICTION ANALYSIS

System / Platform	Data Sources Used	Key Features	Advantages
Traditional Fake News Detection System	Manual news reports, newspaper articles, fact-checking websites	Manual verification, human analysis, static reporting methods	Simple to use, low implementatio n cost, human judgment-based verification
Machine Learning-Based Fake News Detection System	Online news datasets, social media posts, labeled fake and real news articles	Automated classificatio n, pattern recognition, predictive analysis	Faster detection process, improved accuracy, handles large datasets efficiently
AI-Powered Fake News Detection Platform	Real-time online news feeds, digital media platforms, user-generated content	Real-time monitorin g, deep learning models, intelligent prediction systems	High-speed detection, scalable system performance, reduced manual effort
Proposed Fake News Detection System Using NLP	Fake and real news datasets, online news articles, social media text	TF-IDF feature extraction, machine learning classification , automated fake news prediction	High accuracy, real-time analysis, efficient misinformatio n detection, supports decision-making

VI. ALGORITHMIC COMPLIANCE MECHANISMS

The proposed research work titled “Fake News Detection Using Natural Language Processing” uses algorithmic compliance mechanisms to ensure that the fake news detection process is accurate, reliable, transparent, and efficient. In artificial intelligence and machine learning systems, algorithmic compliance refers to the set of methods, rules, and processing techniques used to ensure that the system operates correctly according to predefined objectives and ethical standards. In the context of fake news detection, algorithmic compliance mechanisms help the proposed system analyze textual information systematically, maintain consistency in predictions, reduce classification errors, and improve the reliability of the generated results. These mechanisms also ensure that the machine learning model follows proper data processing procedures and produces fair and meaningful outcomes while analyzing news articles. The proposed system integrates several Natural Language Processing (NLP) and Machine Learning mechanisms to process textual data and classify news articles as fake or genuine. These mechanisms are implemented at different stages of the system, including data preprocessing, feature extraction, model training, classification, and performance evaluation.The first algorithmic compliance mechanism used in the proposed system is data validation and preprocessing compliance. The collected dataset may contain duplicate records, missing values, noisy text, special symbols, punctuation marks, and irrelevant information. If such unwanted data is directly provided to the machine learning model, it can reduce prediction accuracy and create misleading results. Therefore, preprocessing algorithms are applied to validate and clean the dataset before analytical processing begins.

VII. ECONOMIC MODEL FOR ARTIST COMPENSATION

The economic model for artist compensation focuses on ensuring fair, transparent, and sustainable financial support for artists in the digital environment. With the rapid growth of online platforms, streaming services, social media, and digital content distribution systems, artists are now able to reach global audiences more easily than before. However, despite increased visibility and content sharing opportunities, many artists still face challenges related to fair income distribution, copyright protection, unauthorized content usage, and unequal revenue sharing mechanisms. Therefore, an

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 6, Issue 6, June 2026 Impact Factor: 8.2

effective economic compensation model is essential to support artists financially and protect the value of their creative work. The proposed economic model is designed to create a balanced ecosystem where artists receive proper compensation based on the usage, popularity, originality, and audience engagement of their content. The model aims to establish fairness between artists, digital platforms, content distributors, advertisers, and consumers while encouraging creativity and long-term artistic growth.

VIII. ETHICAL CONSIDERATIONS

The proposed research work titled "Fake News Detection Using Natural Language Processing" involves the use of Artificial Intelligence (AI), Machine Learning (ML), and Natural Language Processing (NLP) techniques for analyzing textual information and identifying fake news articles. Since the system processes digital content, user-generated information, and automated decision-making mechanisms, several ethical considerations must be addressed during the design, development, and implementation of the proposed model.

Ethics play an important role in artificial intelligence systems because automated technologies directly influence information accessibility, public opinion, and online communication. A fake news detection system must operate responsibly, fairly, and transparently to avoid misinformation, biased predictions, privacy violations, and misuse of technology. Therefore, ethical principles are integrated into every stage of the proposed research methodology.

One of the primary ethical considerations in this research is data privacy and confidentiality. The proposed system may process large amounts of textual data collected from online news platforms, websites, and social media sources. Some digital content may contain personal information, public opinions, or user-generated comments.

IX. CONCLUSION

The research work titled "Fake News Detection Using Natural Language Processing" focuses on developing an intelligent and automated system capable of identifying fake news articles using Artificial Intelligence (AI), Machine Learning (ML), and Natural Language Processing (NLP) techniques. The rapid growth of digital communication platforms and social media has significantly increased the spread of misinformation across the internet. Fake news not only misleads people but also affects public opinion, social harmony, political systems, healthcare awareness, and overall trust in digital information sources. Therefore, the development of an efficient fake news detection system has become an important requirement in today's digital society.

The proposed research successfully demonstrates how Natural Language Processing techniques can be used to analyze textual content and classify news articles as fake or genuine.

REFERENCES

[1]. Shu, K., Sliva, A., Wang, S., Tang, J., & Liu, H., "Fake News Detection on Social Media: A Data Mining Perspective," ACM SIGKDD Explorations Newsletter, vol. 19, no. 1, pp. 22–36, 2017.
[2]. Ahmed, H., Traore, I., & Saad, S., "Detecting Opinion Spams and Fake News Using Text Classification," Security and Privacy Journal, vol. 1, no. 1, pp. 1–13, 2018.
[3]. Conroy, N. J., Rubin, V. L., & Chen, Y., "Automatic Deception Detection: Methods for Finding Fake News," Proceedings of the Association for Information Science and Technology, vol. 52, no. 1, pp. 1–4, 2015.
[4]. Rubin, V. L., Chen, Y., & Conroy, N. J., "Deception Detection for News: Three Types of Fakes," Proceedings of the Association for Information Science and Technology, vol. 52, no. 1, pp. 1–4, 2015.
[5]. Kaliyar, R. K., Goswami, A., & Narang, P., "FakeBERT: Fake News Detection in Social Media with a BERT-Based Deep Learning Approach," Multimedia Tools and Applications, vol. 80, pp. 11765–11788, 2021.
[6]. Pérez-Rosas, V., Kleinberg, B., Lefevre, A., & Mihalcea, R., "Automatic Detection of Fake News," Proceedings of the 27th International Conference on Computational Linguistics (COLING), pp. 3391–3401, 2018.