An Intelligent Fake News Detection Framework Using Machine Learning and Natural Language Processing

Ankit Raj[1], Ashish Soni[2], Anup Kumar[3], Kabir Kumar[4], Md. Irfan Alam[5]

[1,2,3,4]Students, Bachelor of Computer Applications, Faculty of Computer Science Engineering and Information Technology, Jharkhand Rai University Ranchi, Jharkhand, India

[5]Associate Professor, Faculty of Computer Science Engineering and Information Technology, Jharkhand Rai University Ranchi, Jharkhand, India

[1]mrankit1035@gmail.com, [2]aashishsony7061@gmail.com, [3]anupkumar620478@gmail.com,

[4]kguria982@gmail.com,[5]irfan2.alam2@gmail.com

Abstract

The rapid spread of misinformation on social media and online platforms has become a critical global challenge, impacting public opinion and decision-making in areas such as politics and public health. This paper presents an automated fake news detection system based on machine learning techniques. The proposed approach preprocesses textual data and transforms it into numerical features using the TF-IDF method. Four classification algorithms—Logistic Regression, Naive Bayes, Support Vector Machine (SVM), and Passive Aggressive Classifier—are evaluated for performance. Experimental results demonstrate that Logistic Regression achieves the highest accuracy of approximately 96.8%. Despite its effectiveness, the system is limited by its dependence on training data quality and its inability to detect non-textual misinformation. Overall, the study highlights the potential of machine learning as a practical solution for fake news detection.

Keywords: Fake News Detection, Natural Language Processing, Machine Learning, TF-IDF, Logistic Regression, Text Classification.

1. Introduction

In the digital age, the way people consume news has changed dramatically. Earlier, people mainly relied on newspapers, television, and radio for reliable information. However, with the growth of the internet and social media platforms such as Twitter, Facebook, and WhatsApp, news can now spread instantly to millions of users[2]. While this has made information more accessible, it has also created new challenges. One of the biggest challenges is the rapid spread of fake news. Fake news refers to false, misleading, or fabricated information that is presented in the form of legitimate news. Because social media allows anyone to publish content without strict verification, misinformation can spread quickly and influence a large number of people[1].

The impact of fake news can be serious and far-reaching. It can influence public opinion, affect political decisions, spread fear during crises, and damage the credibility of trustworthy

media organizations. The issue became especially visible during major global events such as the 2016 United States presidential election and the COVID-19 pandemic. During these periods, a large amount of misinformation circulated online, including conspiracy theories, misleading statistics, and false medical advice[3]. Studies have shown that false information spreads faster on social media than accurate information, mainly because sensational or emotional content attracts more attention. As a result, people may unknowingly share misleading content, further increasing the spread of fake news[5].

Because of the enormous amount of content generated online every day, manually verifying each news article is almost impossible. Traditional fact-checking methods require human experts, which makes the process slow and resource-intensive[4]. To solve this problem, researchers have started exploring automated methods using Machine Learning (ML) and Natural Language Processing (NLP). These technologies can analyze large volumes of text data, identify patterns in language, and classify news articles as real or fake. In this project, a fake news detection system is developed using NLP techniques and machine learning algorithms. The system uses TF-IDF for feature extraction and compares different classifiers such as Logistic Regression, Naive Bayes, Support Vector Machine (SVM), and Passive Aggressive Classifier. The goal is to create an efficient and reliable system that can assist in identifying misleading news content automatically.

2. Literature Review

Fake news detection has become an important research area because misinformation spreads quickly on digital platforms. As social media became a major source of news, researchers began developing automated techniques to identify misleading content[6]. Early studies mainly focused on analyzing linguistic patterns such as writing style, word choice, and sentence structure to distinguish fake news from genuine news. Later research expanded this approach by including factors like user behavior, social network patterns, and machine learning models to improve detection accuracy [7, 14].

Several studies have examined different machine learning methods for fake news detection. Ahmad et al. compared algorithms such as Naive Bayes, Support Vector Machine (SVM), and Logistic Regression using a public dataset. Their results showed that Logistic Regression achieved about 94% accuracy when combined with TF-IDF feature extraction [8,9]. Another study by Shu et al. introduced the FakeNewsNet dataset, which considers both textual content and social context such as comments, shares, and user interactions. Although social signals provide valuable insights into how misinformation spreads, collecting such data in real time can be technically difficult [10, 11].

Other research has explored combining multiple signals and analyzing linguistic features of fake news. Ruchansky et al. proposed the CSI deep learning model, which integrates article content, user behavior, and source credibility for detection. Pérez-Rosas et al. found that fake news often contains more emotional, informal, and sensational language than legitimate news. Features such as sentiment, readability, and syntactic complexity can therefore help identify deceptive content. Overall, previous studies show that fake news detection requires

multiple approaches, but for smaller projects a text-based method using TF-IDF and classical machine learning remains a practical and efficient solution.

3. System Architecture

The proposed fake news detection system is a text-based binary classification pipeline that analyzes a news article and classifies it as real or fake. It processes raw news text through multiple stages, including text cleaning, feature extraction, model training, and evaluation, to produce the final result[12,13].

The system architecture mainly includes stages such as text preprocessing, feature extraction, model training, and evaluation. In the preprocessing stage, raw news articles are cleaned by removing punctuation, numbers, HTML tags, and special characters. The text is then converted to lowercase, stop words are removed, and stemming is applied to prepare the data for machine learning analysis.

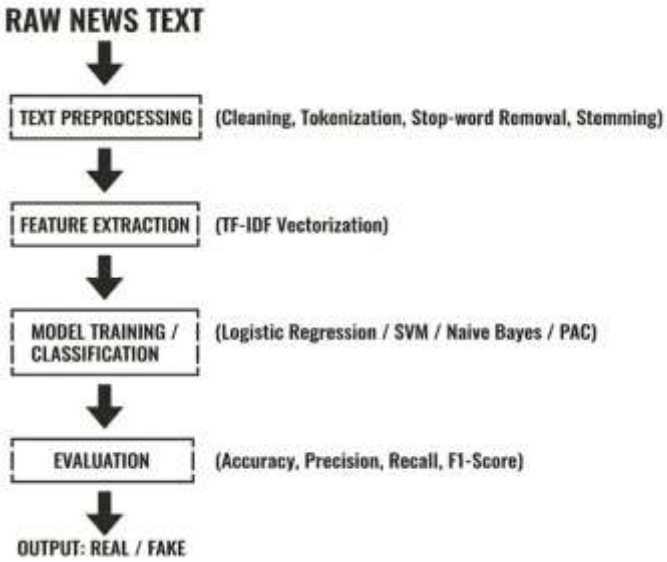

Figure 1. Architecture of the Fake News Detection System.

After preprocessing, feature extraction is performed to convert text into numerical form. In this project, TF-IDF (Term Frequency–Inverse Document Frequency) is used to transform the cleaned text into feature vectors. TF-IDF assigns weights to words based on their importance in a document and across the dataset, helping the system focus on meaningful terms.

Next, machine learning models are trained using these features. The classifiers used include Logistic Regression, Support Vector Machine (SVM), Naive Bayes, and Passive Aggressive Classifier. These models learn patterns from the training data and classify news articles as real or fake.

Finally, the models are evaluated using a separate test dataset. Performance is measured using accuracy, precision, recall, and F1-score. The system follows a modular architecture, allowing components such as feature extraction or classification algorithms to be easily replaced or improved in future developments.

3. Methodology Used

3.1 Dataset Collection

The dataset used in this project is the ISOT Fake News Dataset, developed by the University of Victoria Information Security and Object Technology (ISOT) Research Lab. The dataset contains two CSV files: True.csv and Fake.csv.

The True.csv file contains approximately 21,417 real news articles collected from trusted sources such as Reuters. The Fake.csv file contains about 23,481 fake news articles gathered from unreliable or flagged websites such as GossipCop and Politifact. Each article includes four attributes: title, text, subject, and date.

For this project, the title and text were combined into a single content field to capture more information from the articles. A binary label column was also created where 1 represents real news and 0 represents fake news.

The dataset was divided into training and testing sets using an 80:20 ratio. Stratified sampling was used to maintain a balanced distribution of real and fake news in both sets.

Attribute	Value
Total Articles	44,898
Real News Articles	21,417 (47.7%)
Fake News Articles	23,481 (52.3%)
Average Article Length	~847 words
Training Set (80%)	35,918 articles
Test Set (20%)	8,980 articles
TF-IDF Feature Space	50,000 features
Date Range	2015 – 2018
Real News Source	Reuters.com
Fake News Source	GossipCop, Politifact flagged sites

Figure 2. Dataset Distribution and Train–Test Split

3.2 Text Preprocessing

Raw news text often contains unnecessary elements such as punctuation, numbers, and special characters. Therefore, text preprocessing is applied to clean and prepare the data for machine learning.

First, all text is converted to **lowercase** to maintain consistency. Next, **special characters, numbers, and URLs are removed** using regular expressions. The cleaned text is then **tokenized**, which means splitting the text into individual words.

After tokenization, **stop words** such as the, is, at, and which are removed because they provide little useful information for classification. Finally, **stemming** is applied using the **Porter Stemmer**, which reduces words to their root form. These steps help reduce noise in the dataset and improve model performance.

3.3 Feature Extraction — TF-IDF

Machine learning algorithms cannot directly process text data, so the text must be converted into numerical features. In this project, **TF-IDF (Term Frequency–Inverse Document Frequency)** is used for feature extraction.

TF-IDF assigns weights to words based on how frequently they appear in a document and how rare they are across the entire dataset. Words that appear frequently in a document but rarely across other documents receive higher importance.

The TF-IDF formula is:

$$\textbf{TF-IDF}(t, d) = \textbf{TF}(t, d) \times \log(N / \textbf{DF}(t))$$

where t represents the term, d represents the document, N is the total number of documents, and DF(t) is the number of documents containing the term.

In this project, **TfidfVectorizer from Scikit-learn** was used with a maximum of **50,000 features** and an **n-gram range of (1,2)**.

3.4 Machine Learning Models

Four supervised machine learning algorithms were used to classify news articles as real or fake.

Logistic Regression (LR) is a linear classification model that estimates the probability of a binary outcome. It is widely used in text classification and performs well with TF-IDF features.

Multinomial Naive Bayes (MNB) is based on Bayes' theorem and assumes that features are independent. It is computationally efficient and commonly used for text classification tasks.

Support Vector Machine (SVM) with a linear kernel is used to find the optimal boundary that separates real and fake news in the feature space. It is known for its strong performance in high-dimensional text data.

Passive Aggressive Classifier (PAC) is an online learning algorithm that updates its model when incorrect predictions occur. It works well for large text datasets.

All models were implemented using the Scikit-learn library in Python.

4. Results and Analysis

4.1 Overall Accuracy Comparison

The four machine learning classifiers were evaluated using the same 20% test dataset. The overall accuracy of each model is shown in Table 1.

Table 1: Classifier Accuracy Comparison

Classifier	Accuracy (%)
Logistic Regression	96.82
SVM (Linear Kernel)	96.51
Passive Aggressive Classifier	95.74
Multinomial Naive Bayes	93.18

Logistic Regression achieved the highest accuracy, followed closely by the Linear SVM. Passive Aggressive also performed well, while Multinomial Naive Bayes produced the lowest accuracy

among the four models.

4.2 Detailed Performance Metrics

Detailed evaluation metrics including Precision, Recall, and F1-Score are shown in the following tables.

In the classification reports:

- Class 0 → Fake news
- Class 1 → Real news

Table 2: Logistic Regression — Classification Report

Class	Precision	Recall	F1-Score	Support
Fake (0)	0.97	0.97	0.97	4696
Real (1)	0.97	0.97	0.97	4284
Macro Avg	0.97	0.97	0.97	8980

Table 3: SVM (Linear) — Classification Report

Class	Precision	Recall	F1-Score	Support
Fake (0)	0.97	0.96	0.97	4696
Real (1)	0.96	0.97	0.96	4284
Macro Avg	0.97	0.97	0.97	8980

Table 4: Passive Aggressive Classifier — Classification Report

Class	Precision	Recall	F1-Score	Support
Fake (0)	0.96	0.96	0.96	4696
Real (1)	0.95	0.96	0.96	4284
Macro Avg	0.96	0.96	0.96	8980

Table 5: Multinomial Naive Bayes — Classification Report

Class	Precision	Recall	F1-Score	Support
Fake(0)	0.94	0.94	0.94	4696
Real (1)	0.93	0.92	0.93	4284
Macro Avg	0.93	0.93	0.93	8980

4.3 Confusion Matrix — Logistic Regression

The confusion matrix provides deeper insight into the performance of the Logistic Regression model, which achieved the highest accuracy.

	Predicted Fake	Predicted Real
Actual Fake	4,551 (TN)	145 (FP)
Actual Real	141 (FN)	4,143 (TP)

Explanation:

- True Negatives (4,551): Fake articles correctly identified as fake.
- True Positives (4,143): Real articles correctly identified as real.
- False Positives (145): Fake articles incorrectly classified as real.
- False Negatives (141): Real articles incorrectly classified as fake.

The number of false positives and false negatives is quite similar, indicating that the model does not show a strong bias toward any single class.

4.4 Analysis and Observations (Short & Simple)

All four classifiers performed well, with accuracy ranging from 93% to about 97%. This indicates that TF-IDF features effectively capture important linguistic patterns that help distinguish fake news from real news articles.

Among the models, Logistic Regression achieved the best performance. It produced balanced precision and recall values for both classes, showing that the model does not favor either fake

or real news. This balanced behavior is desirable for datasets where both classes are present in similar proportions.

Although Multinomial Naive Bayes showed the lowest accuracy, it has an important advantage: very fast training time. Because of its simplicity and efficiency, it can be useful in situations where computational resources are limited and quick model training is required.

Another observation is that many misclassified articles were opinion-based or satirical pieces. These articles often use factual language but present information in a misleading or humorous context. Such cases are difficult for text-based machine learning models to classify correctly because the words themselves may appear factual even when the overall meaning is not.

5. Discussion

5.1 Advantages of the Proposed System

The proposed fake news detection system offers several important advantages.

Simplicity and Interpretability:

The system uses Logistic Regression, which is easier to understand compared to complex deep learning models. This makes it possible to identify which words or phrases contribute most to predicting whether an article is fake or real.

Efficiency:

The combination of TF-IDF feature extraction and Logistic Regression is computationally efficient. The model can be trained quickly and does not require powerful hardware, making it suitable for implementation on standard computers.

Strong Baseline Performance:

With an accuracy close to 97%, the system demonstrates strong performance in detecting fake news. This level of accuracy makes it a reliable baseline solution for automated misinformation detection.

Ease of Deployment:

The trained model can be saved using Python libraries such as joblib and easily integrated into applications such as web platforms, news filtering systems, or browser extensions.

5.2 Limitations

Despite its effectiveness, the system also has several limitations.

Dataset Dependency:

The model was trained using the ISOT dataset, which contains specific writing styles and topics. If the system is applied to news from different domains or time periods, its performance may decrease due to changes in language patterns.

Lack of Multimodal Detection:

The current system analyzes text only. However, modern misinformation may include manipulated images, videos, or misleading headlines, which cannot be detected by a text-based model.

No Context or Source Evaluation:
The model classifies articles solely based on textual content. It does not consider factors such as the credibility of the news source, user engagement patterns, or the social context of the information.

Static Model:
Once trained, the model remains unchanged unless it is retrained with new data. Since misinformation strategies evolve over time, periodic updates are necessary to maintain accuracy.

Language Restriction:
The model is trained only on English-language news articles. Extending the system to other languages would require additional datasets and language-specific preprocessing methods.

Difficulty with Satire and Opinion:
Satirical articles and opinion-based content may sometimes be misclassified. These types of content often use realistic language, which can confuse machine learning models that rely primarily on textual patterns.

6. Conclusion and Future Work

This paper presented an end-to-end fake news detection system built using standard NLP and machine learning techniques. The system uses TF-IDF for feature extraction and evaluates four classifiers: Logistic Regression, Naive Bayes, SVM, and Passive Aggressive Classifier. Experimental results on the ISOT Fake News Dataset showed that Logistic Regression achieves the best performance with an accuracy of approximately 96.82%, while maintaining balanced precision and recall across both classes.

The key takeaway is that even a relatively straightforward ML pipeline can achieve high accuracy on text-based fake news detection when trained on a clean, well-labeled dataset. This makes it a viable approach for real-world deployment, particularly in scenarios where computational resources are limited or where model interpretability is important.

Several promising directions can be explored to enhance the proposed system. Advanced deep learning models such as BERT or RoBERTa can be applied to improve accuracy. Incorporating multimodal analysis by combining text with image and video data can help detect visually manipulated content. Developing a real-time web scraping and classification pipeline would increase practical usability. Extending the system to support multiple languages can broaden its impact. Adding explainability techniques like SHAP or LIME can improve transparency and trust. Additionally, integrating social media interaction features

may help identify coordinated misinformation campaigns. Overall, while machine learning cannot fully replace human judgment, it can act as an effective first-level filter in combating fake news.

References

[1] A. Guess, B. Nyhan, and J. Reifler, "Selective exposure to misinformation: Evidence from the consumption of fake news during the 2016 US presidential campaign," European Research Council, vol. 9, 2018.

[2] World Health Organization, "Infodemic management: A key component of the COVID-19 global response," WHO, Geneva, July 2020. Available: https://www.who.int/teams/risk-communication/infodemic-management

[3] S. Vosoughi, D. Roy, and S. Aral, "The spread of true and false news online," Science, vol. 359, no. 6380, pp. 1146–1151, Mar. 2018.

[4] I. Ahmad, M. Yousaf, S. Yousaf, and M. O. Ahmad, "Fake news detection using machine learning ensemble methods," Complexity, vol. 2020, Article ID 8885861, 2020.

[5] K. Shu, A. Sliva, S. Wang, J. Tang, and H. Liu, "Fake news detection on social media: A data mining perspective," ACM SIGKDD Explorations Newsletter, vol. 19, no. 1, pp. 22–36, Sep. 2017.

[6] N. Ruchansky, S. Seo, and Y. Liu, "CSI: A hybrid deep model for fake news detection," in Proc. ACM Int. Conf. Information and Knowledge Management (CIKM), Singapore, Nov. 2017, pp. 797–806.

[7] V. Pérez-Rosas, B. Kleinberg, A. Lefevre, and R. Mihalcea, "Automatic detection of fake news," in Proc. 27th Int. Conf. Computational Linguistics (COLING), Santa Fe, NM, USA, 2018, pp. 3391–3401.

[8] W. Y. Wang, "'Liar, liar pants on fire': A new benchmark dataset for fake news detection," in Proc. 55th Annual Meeting of the Association for Computational Linguistics (ACL), Vancouver, Canada, 2017, pp. 422–426.

[9] A. Kesarwani, S. S. Chauhan, and A. R. Nair, "Fake news detection on social media using K-nearest neighbor classifier," in Proc. Int. Conf. Advanced Computing and Communication Systems (ICACCS), Coimbatore, India, 2020, pp. 189–193.

[10] D. Popat, S. Mukherjee, J. Strötgen, and G. Weikum, "CredEye: A credibility lens for analyzing and explaining misinformation," in Proc. The Web Conference, Lyon, France, 2018, pp. 155–158.

[11] Alam, M. I., & Priya, A. (2025). *Computational approaches to revitalize Maithili literature: Bridging tradition and technology*. In *Recent Advances in Artificial Intelligence for Sustainable Development (RAISD 2025)* (Vol. 196, pp. 245–255). Atlantis Press, Springer Nature.

[12] M. Ahmed, I. Traore, and S. Saad, "Detecting opinion spams and fake news using text classification," Security and Privacy, vol. 1, no. 1, pp. 1–15, Jan. 2018.

[13] F. Monti, F. Frasca, D. Eynard, D. Mannion, and M. M. Bronstein, "Fake news detection on social media using geometric deep learning," arXiv preprint, arXiv:1902.06673, Feb. 2019

[14] B. D. Horne and S. Adah, "This just in: Fake news packs a lot in title, uses simpler, repetitive content in text body, more similar to satire than real news," in Proc. Int. Workshop on News and Public Opinion (ICWSM), 2017.