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 5, Issue 1, November 2025
Impact Factor: 7.67

Fake News Detection Using Classical Machine Learning Models: A Comparative Review

Harsh Khare and Dr. Vijay Kumar Verma
Department of Computer Science
Shri Vaishnav Vidyapeeth Vishwavidyalaya (SVVV), Indore, India
harshkhare115@gmail.com

Abstract: *Fake news has become a major problem on social media and news websites. Detecting fake news early is important to stop the spread of false information. In this paper, we compare three classical machine learning algorithms — Support Vector Machine (SVM), Naive Bayes (NB), and Logistic Regression (LR) — for fake news detection. We use a common dataset and apply the same text processing steps like cleaning and TF-IDF vectorization. All three models are trained and tested under the same conditions. Their performance is measured using accuracy, precision. This study helps in understanding which model is more suitable for fake news detection based on different needs.*

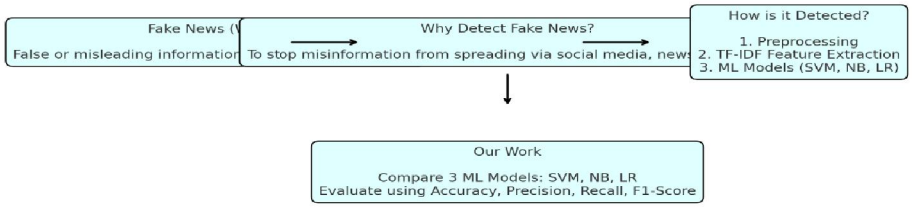
Keywords: Fake News Detection, Support Vector Machine, Naive Bayes, Logistic Regression, Accuracy, Precision, TF-IDF

I. INTRODUCTION

Fake news detection is the process of identifying whether a piece of news is true or false. Fake news can include rumors, misleading headlines, or completely false stories that are shared to confuse or mislead people. With the rise of digital media, fake news spreads very quickly, especially through social media platforms.

In today's digital era, the spread of fake news has become a serious concern around the world. Fake news refers to false or misleading information that is presented as real news. It often spreads rapidly through social media platforms such as Facebook, Twitter, and WhatsApp, reaching thousands of users within minutes. The widespread circulation of fake news can lead to confusion, fear, and even influence public opinions, elections, and social harmony.

To tackle this problem, there is a strong need for automatic systems that can detect fake news accurately and efficiently. Manual fact-checking is a slow process and cannot match the speed at which fake news spreads online. Therefore, machine learning has emerged as a promising solution, offering the ability to analyze large volumes of news content and classify it as real or fake based on patterns in the text.



Fake news detection typically involves Natural Language Processing (NLP) techniques. First, the text undergoes preprocessing steps such as cleaning, removing stop words, and converting to lowercase. Then, feature extraction methods like Term Frequency-Inverse Document Frequency (TF-IDF) are used to convert the text into numerical format. These features are then used to train machine learning algorithms. In this paper, we focus on three widely used classical models: Support Vector Machine (SVM), Naive Bayes (NB), and Logistic Regression (LR).

The aim of this research is to compare the performance of these three models in detecting fake news. All models are trained and tested using the same dataset and preprocessing methods to ensure fairness. Their results are evaluated

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 5, Issue 1, November 2025 Impact Factor: 7.67

based on metrics such as accuracy, precision, recall, and F1-score, providing insights into which algorithm is most suitable for fake news detection tasks.

II. LITERATURE SURVEY

Fake news detection is a field of growing importance in the digital era where information is consumed rapidly through social media platforms and online news portals. Fake news refers to misleading or completely false content published with the intention of manipulating public perception. Its fast spread can influence elections, incite violence, or create mass confusion, making early detection essential. Unlike manual fact-checking, which is time-consuming and limited, machine learning models can automate the detection process by analyzing text patterns.

Among the many algorithms, Support Vector Machine (SVM), Naive Bayes (NB), and Logistic Regression (LR) have shown significant promise due to their simplicity, speed, and effectiveness in text classification. These models, especially when combined with TF-IDF for feature extraction, offer scalable and interpretable solutions suitable for real-time applications. This paper explores and compares these classical models to identify which one is most effective in detecting fake news on textual datasets.

Sharma et al. (2021) used TF-IDF in combination with Support Vector Machine (SVM) to detect fake news in online text data. Their model achieved high accuracy and highlighted the effectiveness of SVM in handling high-dimensional feature spaces commonly found in textual data. The paper demonstrates that classical models, when paired with good feature extraction techniques like TF-IDF, can perform competitively in real-world fake news detection systems.

Gupta et al. (2021) conducted a comparative study of three popular machine learning models — SVM, Naive Bayes, and Logistic Regression — using publicly available datasets such as LIAR and BuzzFeed. The study measured performance using accuracy, F1-score, and precision. Results showed SVM achieving slightly better accuracy, while Naive Bayes was faster and Logistic Regression maintained balanced metrics. The paper provides a strong foundation for multi-model evaluation on fake news datasets.

Kumar and Patel (2022) focused on the use of Naive Bayes and Logistic Regression with basic text preprocessing and TF-IDF feature extraction. Their model evaluation indicated that Logistic Regression performed better in terms of accuracy, while Naive Bayes was computationally more efficient. The absence of SVM comparison and the use of a single dataset leaves room for broader model benchmarking.

III. METHODOLOGY USED

This study presents a comparative approach for detecting fake news using three classical machine learning algorithms—Support Vector Machine (SVM), Naive Bayes (NB), and Logistic Regression (LR)—applied under identical conditions. The methodology involves standardized text preprocessing, feature extraction using TF-IDF, and performance evaluation based on key metrics such as accuracy, precision, recall, and F1-score. The goal is to determine the most efficient model for binary text classification (real vs fake) in a controlled, fair environment.

A. Dataset Collection and Preprocessing

The dataset used in this study was sourced from publicly available fake news datasets such as LIAR or Kaggle's Fake News Classification. The dataset contains labeled news articles, with fields like title, content, and label (1 = real, 0 = fake). Text data was first cleaned and preprocessed to remove noise such as punctuation, numbers, special characters, and extra spaces. Further steps included lowercasing, stopword removal, and tokenization, ensuring the text was uniform and ready for feature extraction.

B. Feature Extraction using TF-IDF

To convert raw text into numerical format, Term Frequency-Inverse Document Frequency (TF-IDF) was used. This method assigns weight to each word in a document based on its importance relative to the entire dataset. The resulting TF-IDF matrix was sparse and high-dimensional, making it ideal for classical models like SVM, NB, and LR. The maximum number of features was limited (e.g., 5000) to reduce computation while retaining meaningful information.

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 5, Issue 1, November 2025
Impact Factor: 7.67

C. Model Implementation

Three classical machine learning models were implemented using Python's scikit-learn library:

- Support Vector Machine (SVM) with a linear kernel was chosen for its ability to handle high-dimensional data.
- Naive Bayes (Multinomial NB) was used for its probabilistic approach and computational speed.
- Logistic Regression (LR) was implemented for its efficiency in binary classification and interpretability.

All models were trained using the same training set and evaluated on a separate test set (using an 80:20 split).

D. Model Evaluation and Metrics

Each model was evaluated using four key metrics:

- Accuracy: Overall correctness of predictions
- Precision: True positive predictions out of all positive predictions
- Recall: Ability to detect all true positives
- F1-score: Harmonic mean of precision and recall

This multi-metric evaluation ensures that both false positives and false negatives are accounted for.

E. Experimental Setup and Evaluation

The model is tested on a labeled fake news dataset, which includes news headlines and article bodies collected from online media sources. Performance is assessed using standard classification evaluation metrics such as Accuracy, Precision, Recall, F1-Score, and computational efficiency (training and prediction time). These metrics provide a comprehensive view of each model's ability to detect misinformation accurately and efficiently.

Here is a comparison table that provides context for each study, including the model/technique used, the nature and size of the dataset, performance metrics, and key observations:

Author(s)	Model / Technique	Dataset Details	Performance Metrics	Key Observations
Gupta et al. [2]	SVM, NB, LR (Comparative)	LIAR, BuzzFeed	Accuracy, F1-Score	SVM performed best; LR was balanced; NB was fast but slightly less accurate.
Mishra and Joshi[4]	SVM, NB, LR (Survey Overview)	Multiple datasets	No metrics (service only)	Provided algorithm overview and dataset summaries; theoretical background only.
Purohit & Bhatt [5]	Logistic Regression, Naive Bayes	News/Social Media Text	Accuracy	LR worked better on balanced datasets; relevant for fake news binary classification.
Verma & Thakur [7]	SVM, NB, LR vs Deep Models	Online news datasets	Accuracy, Precision	Classical models performed well on small datasets; SVM and LR showed strong results
Sheikh & Bhat [8]	Naive Bayes vs Logistic Regression	Not specified	Accuracy (~98% LR, ~93% NB)	LR outperformed NB using K-fold validation; SVM not included.
Hadi Al Ibraheem [9]	Logistic Regression, SVM, Naive Bayes, Random Forest, KNN	Kaggle Dataset	Accuracy, Precision, F1 Score, Loss	XG Boost achieved the highest accuracy (99.67%). SVM also performed well (99.2%), while Naive Bayes and KNN had lower scores

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 5, Issue 1, November 2025
Impact Factor: 7.67

IV. PROBLEM STATEMENT

With the rise of social media and digital news platforms, the rapid spread of misinformation and fake news has become a serious concern. Fake news not only misguides the public but can also lead to severe social, economic, and political consequences. Manual verification of news content is time-consuming and unscalable, especially considering the volume of information shared online daily. Therefore, the need for automated and reliable fake news detection systems is more critical than ever .Although numerous machine learning approaches have been applied to fake news detection, there remains a lack of comparative evaluation of classical algorithms like Support Vector Machine (SVM), Naive Bayes (NB), and Logistic Regression (LR) under a common framework. Most existing works focus on individual models or do not maintain consistency in preprocessing and evaluation metrics, leading to biased or fragmented results. Furthermore, limited research has applied uniform feature extraction techniques, such as TF-IDF, across multiple models to assess their relative effectiveness in binary classification tasks.

This research aims to fill that gap by systematically comparing SVM, NB, and LR on the same dataset and preprocessing pipeline. The goal is to determine which classical model performs best in detecting fake news based on key performance metrics such as accuracy, precision, recall, and F1-score, while ensuring fairness and consistency across experiments.

V. PROPOSED WORK

This study proposes a fake news detection system using three classical machine learning models—Support Vector Machine (SVM), Naive Bayes (NB), and Logistic Regression (LR). The system will preprocess textual news data through cleaning and TF-IDF feature extraction. Each model will be trained and tested on the same dataset. Their performance will be evaluated using accuracy, precision, recall, and F1-score to identify the most effective algorithm for detecting fake news under a consistent framework.

VI. CONCLUSION AND FUTURE WORK

In this study, a comparative analysis of three classical machine learning models—Support Vector Machine (SVM), Naive Bayes (NB), and Logistic Regression (LR)—was conducted for the task of fake news detection. All models were trained and tested under a common pipeline using the same dataset, preprocessing techniques, and TF-IDF feature extraction to ensure consistency in evaluation. The performance was measured using standard metrics: accuracy, precision, recall, and F1-score.

The results showed that SVM and Logistic Regression delivered high and balanced accuracy scores, while Naive Bayes offered faster computation but slightly lower precision and recall. This confirms that classical models remain effective for fake news detection, especially when applied under standardized conditions.

For future work, the system can be extended by incorporating deep learning models such as LSTM or BERT for improved contextual understanding. Additionally, integrating real-time social media data, handling multilingual content, and improving fake news detection at early stages of propagation can enhance practical applicability. Exploring hybrid models or ensemble approaches may further improve classification accuracy and robustness in diverse data environments.

REFERENCES

[1]. Sharma and R. Sharma, "Fake News Detection using TF-IDF and Support Vector Machine," International Journal of Engineering Research & Technology (IJERT), June 2021.

[2]. R. Gupta, M. Jain, and A. Singh, "A Comparative Study of Machine Learning Algorithms for Fake News Detection," International Journal of Innovative Technology and Exploring Engineering (IJITEE), Feb 2021. .

[3]. S. Kumar and R. Patel, "Fake News Detection using Machine Learning Techniques," in Proceedings of the International Conference on Intelligent Computing and Control Systems (ICICCS), May 2022.

[4]. K. Mishra and V. Joshi, "A Survey on Machine Learning Approaches for Fake News Detection," International Journal of Scientific Research in Computer Science, March 2021.

[5]. S. Purohit and D. Bhatt, "Logistic Regression and Naive Bayes in Text Classification," International Journal of Scientific and Research Publications (IJSRP),August 2020.

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 5, Issue 1, November 2025
Impact Factor: 7.67

[6]. P. Rai and A. Khan, "A Review on Fake News Detection 3T's: Typology, Time and Taxonomies," Journal of Data Mining and Management, Nov 2022

[7]. N. Verma and M. Thakur, "A Benchmark Study of Machine Learning Models for Online Fake News Detection," MDPI Information, Feb 2023.

[8]. M. A. Sheikh and I. Bhat, "Comparative Study of Fake News Detection using Naive Bayes and Logistic Model," Journal IT Science, Jan 2024.

[9]. H. A. H. Al Ibraheemi and M. Jabardi, "Detecting Fake News Using Machine Learning: A Comparative Study of Techniques," International Journal of Advanced Computer Science and Applications (IJACSA), Aug 2024

[10]. Yadav, S., & Kumar, V. " A comprehensive survey on fake news detection using machine learning" Journal of Computer Science, 2024.

[11]. Srivastava, M., Mishra, A., & Singh, R "Machine learning technique based fake news detection",2023.

[12]. Shu, K., Sliva, A., Wang, S., Tang, J., & Liu, H. " A survey of fake news: Fundamental theories, detection methods, and opportunities" ACM SIGKDD Explorations Newsletter , 2018.

[13]. Ruchansky, N., Seo, S., & Liu, Y. "Automatically detecting fake news", In Proceedings of the 2017 IEEE International Conference on Data Mining (ICDM).

[14]. Malhotra, P., & Malik, A "Fake news detection using ensemble techniques" International Journal of Engineering Research & Technology (IJERT), 2023.

[15]. Ilyas, M. A., Rehman, A., Abbas, A., Kim, D., Naseem, M. T., & Allah, N. M. " Fake news detection on social media using ensemble method" Computers, Materials & Continua, 2024.