iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 14 **Issue:** I **Month of publication:** January 2026

DOI: https://doi.org/10.22214/ijraset.2026.77015

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue I Jan 2026- Available at www.ijraset.com

Fake News Detection Using Machine Learning

Sonali Dinesh Chawre
Department of Computer Science and Engineering, Guru Nanak Institute of Engineering & Technology, Nagpur, India

Abstract: ***The exponential growth of digital media and social networking platforms has transformed the way information is created and consumed. While these platforms provide rapid access to news, they have also facilitated the widespread dissemination of fake news. Fake news refers to intentionally false or misleading information presented as authentic news, which can influence public opinion, disrupt social harmony, and erode trust in media. Traditional manual fact-checking approaches are time-consuming and inadequate for handling the massive volume of online content. Therefore, automated fake news detection systems are essential.***

This paper proposes a machine learning-based approach for detecting fake news using Natural Language Processing (NLP) techniques. The proposed system pre-processes textual data and applies feature extraction methods such as Bag of Words (BoW) and Term Frequency–Inverse Document Frequency (TF-IDF). Multiple machine learning classifiers, including Logistic Regression, Naïve Bayes, Random Forest, and Support Vector Machine, are trained and evaluated. The performance of these models is assessed using accuracy, precision, recall, and F1-score. Experimental results demonstrate that classical machine learning models can effectively classify news articles as fake or real with reliable accuracy. The proposed system offers a lightweight and efficient solution for fake news detection and can be extended to real-time applications in the future.

Keywords: Fake News, Machine Learning, Natural Language Processing, TF-IDF, Bag of Words, Classification

I. INTRODUCTION

The rapid advancement of the internet and social media platforms has revolutionized the dissemination of information. Online news portals, blogs, and social networking sites enable users to access information instantly. However, this convenience has also led to the uncontrolled spread of fake news. Fake news is defined as false or misleading information deliberately created and circulated to deceive readers. It has emerged as a major societal issue due to its influence on public opinion, political processes, and social stability.

The impact of fake news can be severe. During elections, misleading information can manipulate voter behavior, while during health crises such as pandemics, misinformation can lead to panic and unsafe practices. Traditional fact-checking techniques rely on human experts and journalists, making them slow and ineffective for managing large-scale data generated on digital platforms.

Recent developments in Machine Learning (ML) and Natural Language Processing (NLP) provide promising solutions to this challenge. By analyzing linguistic patterns, writing styles, and statistical features of textual data, ML models can automatically distinguish between fake and real news. Automated fake news detection systems can process vast amounts of data efficiently and provide timely predictions.

This paper focuses on developing a machine learning-based fake news detection system using classical ML algorithms. The proposed approach involves preprocessing textual data, extracting meaningful features, and evaluating multiple classifiers to identify the most effective model. The objective is to design an accurate and efficient system suitable for academic research and practical applications.

II. LITERATURE REVIEW

Fake news detection has gained significant research attention in recent years. Several studies have explored different approaches, ranging from traditional machine learning techniques to advanced deep learning models.

Shu et al. (2017) proposed a data mining framework for fake news detection that integrates content-based and social context-based features. Their work highlighted the importance of analysing both news content and its propagation patterns on social media. However, the framework lacked real-time detection capabilities.

Ahmed et al. (2017) investigated the use of N-gram analysis combined with machine learning classifiers such as Naïve Bayes and Support Vector Machines. Their results demonstrated that classical ML models could achieve high accuracy with relatively low computational complexity.

Ruchansky et al. (2017) introduced the CSI model, which combined text analysis with user behaviour modelling using deep learning techniques. Although the model achieved improved accuracy, it required large-scale datasets and high computational resources.
Kaliyar et al. (2020) applied deep learning models such as LSTM and CNN for fake news detection and reported enhanced performance compared to traditional methods. However, the complexity and training time of deep learning models make them less suitable for lightweight applications.
From the literature, it is evident that while deep learning methods offer high accuracy, classical machine learning techniques remain practical due to their simplicity, interpretability, and lower computational cost. This research focuses on evaluating classical ML algorithms for effective fake news detection.

III. PROPOSED METHODOLOGY

The proposed fake news detection system follows a systematic workflow consisting of data collection, pre-processing, feature extraction, model training, and evaluation.

A. Dataset Description

A publicly available benchmark dataset, such as the Kaggle Fake News dataset, is used for experimentation. The dataset contains labelled news articles classified as fake or real, providing a reliable basis for supervised learning.

B. Data Pre-processing

Text pre-processing is a crucial step in NLP-based systems. The pre-processing process includes:

- Conversion of text to lowercase
- Removal of punctuation and special characters
- Tokenization of text
- Removal of stop words
- Stemming and lemmatization

These steps help reduce noise and standardize the textual data.

C. Feature Extraction

To convert textual data into numerical form, feature extraction techniques such as Bag of Words (BoW) and TF-IDF are employed. These methods represent text based on word frequency and importance, enabling effective input to machine learning models.

D. Machine Learning Models

The following classifiers are trained and evaluated:

- Logistic Regression
- Naïve Bayes
- Random Forest
- Support Vector Machine (SVM)

Each model is trained using the extracted features and optimized for classification performance.

IV. EXPERIMENTAL SETUP AND RESULTS

The dataset is divided into training and testing sets using an 80:20 split. Model performance is evaluated using accuracy, precision, recall, and F1-score.

A. Performance Metrics

- Accuracy: Overall correctness of the model
- Precision: Proportion of correctly predicted fake news
- Recall: Ability to identify actual fake news
- F1-Score: Harmonic mean of precision and recall

B. Results

TABLE I
COMPARATIVE PERFORMANCE ANALYSIS OF ML CLASSIFIERS

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Logistic Regression	89.2	88.5	87.9	88.2
Naïve Bayes	86.4	85.1	84.7	84.9
Random Forest	91.0	90.2	89.6	89.9
Support Vector Machine	90.1	89.4	88.8	89.1

The results indicate that Random Forest and SVM classifiers outperform other models in terms of overall accuracy, while Logistic Regression also provides competitive performance with lower computational complexity.

V. TOOLS AND TECHNOLOGIES

The following tools and technologies are used:

- Programming Language: Python
- Libraries: Scikit-learn, NLTK, Pandas, Numpy
- Development Environment: Jupyter Notebook / Google Colab
- Visualization: Matplotlib

These tools are widely used in the research community and support reproducibility.

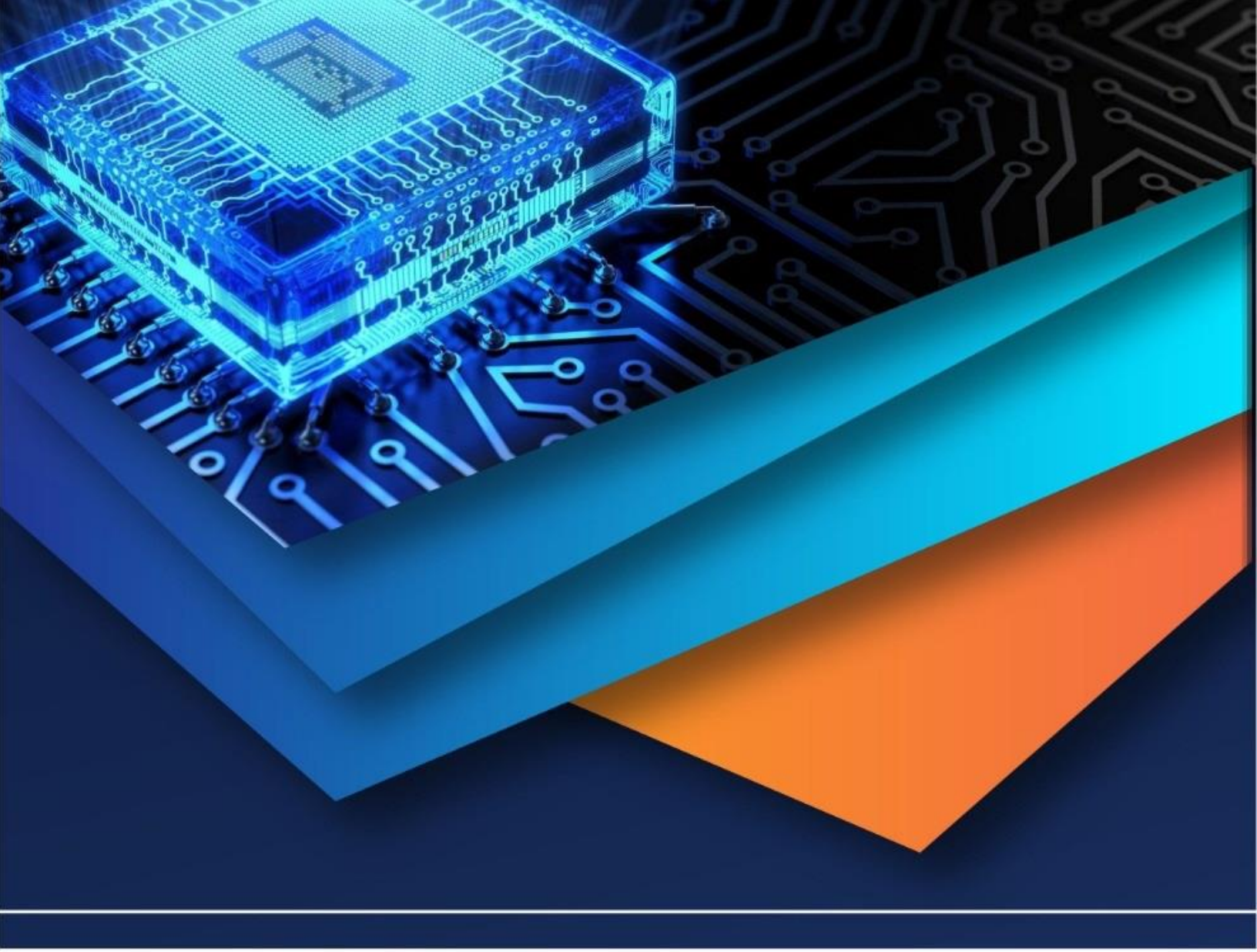
VI. CONCLUSION AND FUTURE SCOPE

This paper presented a machine learning-based approach for fake news detection using NLP techniques. Classical machine learning models were trained and evaluated using TF-IDF and BoW features. Experimental results showed that Random Forest and SVM achieved the highest accuracy, while Logistic Regression offered a lightweight and efficient alternative.

The proposed system demonstrates that classical ML techniques are effective for fake news detection and suitable for academic and real-world applications. In the future, this work can be extended by incorporating deep learning models such as LSTM and BERT, using multilingual datasets, and deploying the system as a real-time web or mobile application.

REFERENCES

[1] 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, 2017.
[2] H. Ahmed, I. Traore, and S. Saad, “Detection of Online Fake News Using N-Gram Analysis and Machine Learning Techniques,” in Proc. Int. Conf. Intelligent, Secure and Dependable Systems, Springer, 2017.
[3] N. Ruchansky, S. Seo, and Y. Liu, “CSI: A Hybrid Deep Model for Fake News Detection,” in Proc. ACM CIKM, 2017.
[4] R. K. Kaliyar, A. Goswami, and P. Narang, “DeepFake: Improving Fake News Detection Using Deep Learning,” IEEE Access, vol. 8, 2020.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY