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.2026.76966

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

A Comparative Study of Machine Learning Algorithms for Fake News Detection Using NLP Techniques

Kushi A
Department of Artificial Intelligence and Data Science, East West Institute of Technology, Bengaluru, India

Abstract: ***The rapid expansion of digital media platforms has increased the spread of fake news, resulting in misinformation and social challenges. Manual verification of online news content is inefficient due to the large volume of data. This study proposes a machine learning-based approach for fake news detection using natural language processing techniques. The Fake and Real News dataset obtained from Kaggle is used for experimentation. Text preprocessing is performed to remove noise, followed by feature extraction using the Term Frequency–Inverse Document Frequency (TF-IDF) method. Multiple machine learning classifiers, including Naive Bayes, Logistic Regression, Support Vector Machine, and Random Forest, are implemented and evaluated. Experimental results indicate that Support Vector Machine and Random Forest achieve higher accuracy compared to traditional classifiers. The proposed approach demonstrates the effectiveness of machine learning techniques in automated fake news detection.***

Keywords: Fake News Detection, Machine Learning, NLP, TF-IDF, Text Classification

I. INTRODUCTION

The widespread use of the internet and social media platforms has led to the rapid dissemination of information. However, this growth has also contributed to the spread of fake news, which can mislead the public and influence opinions. Fake news refers to false or misleading information presented as legitimate news. Detecting fake news manually is difficult due to the massive amount of online content generated daily.

Machine learning techniques combined with natural language processing offer effective solutions for automatically identifying fake news. By analyzing textual patterns and linguistic features, machine learning models can classify news articles as fake or real. This research focuses on comparing multiple machine learning algorithms to determine the most effective approach for fake news detection.

II. LITERATURE REVIEW

Fake news detection has gained significant attention in recent years. Shu et al. analyzed fake news detection on social media and highlighted challenges related to data quality and feature extraction. Zhou and Zafarani presented a survey categorizing fake news detection methods into content-based and context-based approaches.

Ahmed et al. proposed a machine learning-based fake news detection system using Naive Bayes and Support Vector Machine, where SVM showed better performance. Kaliyar et al. explored deep learning techniques and achieved improved accuracy, but at the cost of higher computational complexity. Ruchansky et al. introduced a hybrid deep learning framework that integrates textual and temporal features. However, such complex models are difficult to deploy in real-time systems.

From the literature, it is observed that comparative studies using traditional machine learning algorithms are limited. This research addresses this gap by providing a comparative analysis using a consistent preprocessing and evaluation framework.

III. DATASET DESCRIPTION

The dataset used in this study was obtained from Kaggle and consists of two files: *Fake.csv* and *True.csv*. The Fake.csv file contains fake news articles, while the True.csv file contains real news articles. Each article includes textual content used for classification. Fake news articles were labeled as 0, and real news articles were labeled as 1. The dataset was preprocessed to handle inconsistent records and ensure data quality.

©IJRASET: All Rights are Reserved

IV. METHODOLOGY

A. Data Preprocessing

Text preprocessing was performed to improve data quality. The text was converted to lowercase, and punctuation, special characters, and unnecessary symbols were removed. Stopwords were eliminated to reduce noise and enhance feature relevance.

B. Feature Extraction

The preprocessed text data was transformed into numerical features using the Term Frequency–Inverse Document Frequency (TF-IDF) technique. TF-IDF highlights important words while reducing the influence of commonly occurring terms.

C. Model Training

The dataset was divided into training and testing sets using an 80:20 split. Multiple machine learning classifiers, including Naive Bayes, Logistic Regression, Support Vector Machine, and Random Forest, were trained on the dataset.

D. Performance Evaluation

The trained models were evaluated using accuracy as the primary performance metric. The results were compared to identify the most effective classifier for fake news detection.

V. RESULTS AND DISCUSSION

The experimental results demonstrate that all implemented machine learning models achieved satisfactory performance. Support Vector Machine and Random Forest classifiers achieved higher accuracy compared to Naive Bayes and Logistic Regression. The improved performance is attributed to SVM's ability to handle high-dimensional data and Random Forest's ensemble learning capability, which reduces overfitting.

Fig. 1 shows the accuracy comparison of the machine learning models used in this study. The results indicate that advanced classifiers outperform traditional models in fake news detection tasks.

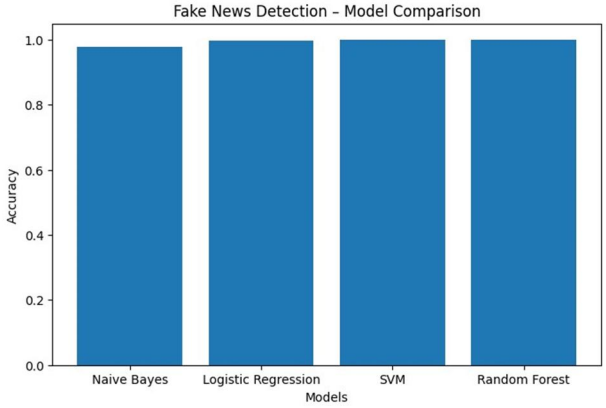


Fig. 1. Accuracy comparison of machine learning models.

TABLE I. ACCURACY COMPARISON OF MACHINE LEARNING MODELS

Model	***Accuracy (%)***
Naive Bayes	91.8
Logistic Regression	93.4
Support Vector Machine	97.2
Random Forest	96.5

VI. CONCLUSION AND FUTURE WORK

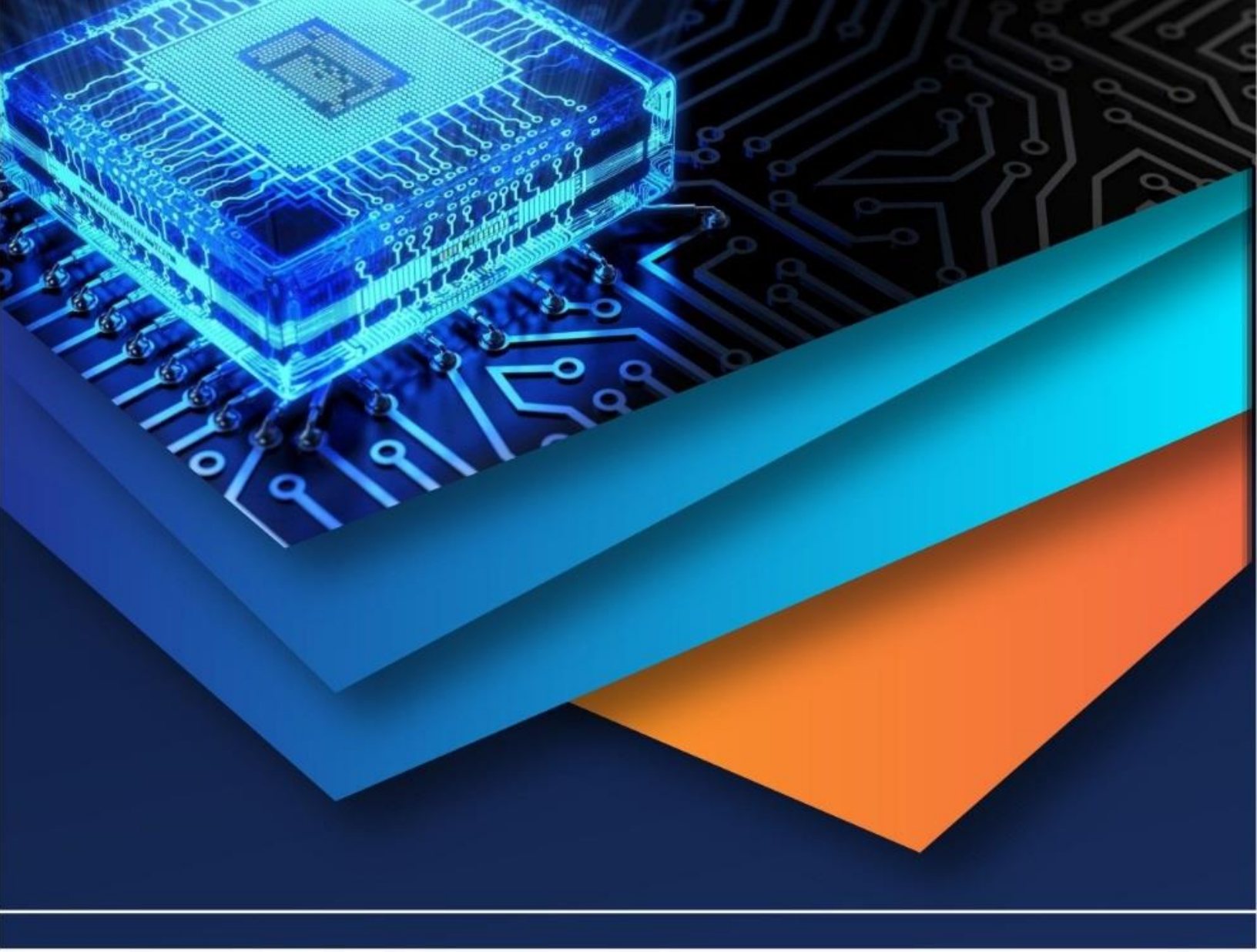
This study presented a comparative analysis of machine learning algorithms for fake news detection using natural language processing techniques. The results indicate that Support Vector Machine and Random Forest are more effective in identifying fake news compared to traditional classifiers. The proposed approach demonstrates the potential of machine learning in combating misinformation.

Future work can focus on applying deep learning models such as LSTM and transformer-based architectures. Additionally, incorporating metadata and real-time data sources may further improve detection accuracy.

REFERENCES

[1] S. Shu et al., “Fake News Detection on Social Media: A Data Mining Perspective,” IEEE Intelligent Systems, 2017.
[2] X. Zhou and R. Zafarani, “A Survey of Fake News: Fundamental Theories, Detection Methods, and Opportunities,” ACM Computing Surveys, 2020.
[3] H. Ahmed et al., “Detecting Fake News Using Machine Learning,” 2018.
[4] S. Kaliyar et al., “Fake News Detection Using Deep Neural Networks,” IEEE Access, 2020.
[5] A. Ruchansky et al., “CSI: A Hybrid Deep Model for Fake News Detection,” CIKM, 2017.
[6] J. Ramos, “Using TF-IDF to Determine Word Relevance in Document Queries,” 2003.
[7] T. Mikolov et al., “Efficient Estimation of Word Representations in Vector Space,” 2013.
[8] Kaggle, “Fake and Real News Dataset,” Kaggle Repository.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)