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 12, May 2026
Impact Factor: 8.2

Fake News Detection using Machine Learning Algorithms

Gitanjali Jatin Thakur[1], Rizwana Momin[2], Soni Pandey[3], Shivani Sonkar[4], Trupti Kasare[5]
Assistant Professor, Department of Computer Science[1-5]
Ramsheth Thakur College of Commerce Science, Kharghar, India

Abstract: *The rapid expansion of digital communication platforms has significantly increased the spread of misinformation and fabricated news stories. Fake news has the potential to influence political opinions, public health decisions, and social stability [1]. Detecting fake news manually is challenging because of the enormous volume of online information generated daily. Therefore, automated detection techniques based on Machine Learning have gained significant attention in recent years [2].*
This research proposes a machine learning–based approach for identifying fake news by analysing textual features extracted from news articles. Several classification algorithms are applied to distinguish between real and fake news articles. Experimental results show that machine learning models can effectively classify news content with high accuracy [3].

Keywords: Fake News Detection, Machine Learning, Text Classification, Natural Language Processing, Data Mining.

I. INTRODUCTION

The digital revolution has transformed the way information is produced, shared, and consumed. Online platforms allow individuals to access news instantly and share information with a global audience. However, this accessibility has also created an environment where false or misleading information spreads rapidly. Fake news refers to deliberately fabricated or manipulated information that appears credible but is intended to deceive readers.

Social media platforms and online news portals enable rapid information distribution, making it difficult to verify the authenticity of content before it reaches a large audience. The consequences of fake news can be significant, including political manipulation, public panic, and damage to institutional credibility.

Traditional fact-checking approaches rely on human experts who manually verify information. Although accurate, this process is time-consuming and cannot keep up with the speed at which misinformation spreads online. As a result, automated solutions using machine learning techniques have emerged as an effective approach for detecting fake news.

This study aims to develop a machine learning-based system capable of identifying fake news by analysing textual patterns and linguistic features present in news articles.

II. LITERATURE REVIEW

Numerous studies have explored automated methods for detecting misinformation using computational techniques. Early research primarily focused on analysing linguistic features and statistical patterns within news articles.

Several researchers have applied supervised machine learning algorithms for text classification tasks. Algorithms such as Naïve Bayes, Support Vector Machines, Logistic Regression, and Decision Trees have demonstrated promising results in distinguishing between authentic and misleading content.

Recent studies have also incorporated advanced natural language processing methods to analyse semantic meaning, sentiment, and writing style. These approaches allow models to capture deeper relationships within textual data.

Despite the progress in this field, fake news detection remains challenging because deceptive content often mimics legitimate journalistic writing styles. Therefore, continuous improvements in data preprocessing, feature extraction, and model training are necessary to enhance detection accuracy.

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 12, May 2026
Impact Factor: 8.2

III. PROBLEM STATEMENT

The rapid spread of misinformation through online platforms has become a serious concern in modern society. Since digital platforms allow users to publish and share information without strict verification mechanisms, fake news can easily circulate among large audiences.

Manual verification methods are insufficient to handle the massive volume of information generated daily. Consequently, there is a need for automated systems capable of identifying misleading information efficiently and accurately.

The primary problem addressed in this research is the development of an automated fake news detection system that can classify news articles as genuine or fake using machine learning techniques.

IV. OBJECTIVES OF THE STUDY

The major objectives of this research include:

1. To examine the characteristics of fake and genuine news articles.
2. To preprocess textual data and prepare it for machine learning models.
3. To extract meaningful features from news content using natural language processing techniques.
4. To implement and compare different machine learning algorithms for classification.
5. To evaluate the effectiveness of the proposed system using standard performance metrics.

V. METHODOLOGY

5.1 Data Collection

The dataset used for this research consists of news articles labelled as either fake or real. Each record contains textual content along with classification labels that indicate the authenticity of the news article.

The dataset is divided into training and testing sets to evaluate the performance of the machine learning models.

5.2 Data Preprocessing

Raw textual data typically contains noise and unnecessary information that can reduce model performance. Therefore, several preprocessing steps are applied to clean and standardize the dataset.

The preprocessing stage includes:

Converting text to lowercase

Removing punctuation and special characters

Eliminating commonly used stop words

Tokenizing sentences into individual words

Applying stemming or lemmatization techniques

These steps help improve the quality of textual features used during model training.

5.3 Feature Extraction

Machine learning models require numerical input, therefore textual data must be transformed into numerical representations. Feature extraction techniques are used to convert text into meaningful vectors.

Commonly used feature extraction methods include:

Bag of Words (BoW) representation

Term Frequency–Inverse Document Frequency (TF-IDF)

Word frequency analysis

These techniques capture the importance of words and phrases within news articles, enabling the classification algorithms to identify patterns associated with fake or real news.

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 12, May 2026 Impact Factor: 8.2

5.4 Machine Learning Models

Several supervised machine learning algorithms are implemented to classify news articles. These include:

Logistic Regression

A statistical classification technique used for predicting binary outcomes.

Decision Tree

A hierarchical model that classifies data by splitting it into branches based on decision rules.

Random Forest

An ensemble learning method that combines multiple decision trees to improve classification accuracy and reduce overfitting.

Each algorithm is trained using the extracted textual features and evaluated based on classification performance.

VI. PROPOSED SYSTEM ARCHITECTURE

The proposed fake news detection system consists of the following stages:

1. Dataset collection
2. Text preprocessing
3. Feature extraction
4. Model training
5. Model evaluation
6. News classification

This pipeline ensures that raw news data is processed systematically before being analysed by machine learning algorithms.

VII. EXPERIMENTAL RESULTS AND ANALYSIS

The performance of the machine learning models is evaluated using standard classification metrics including accuracy, precision, recall, and F1 score.

Experimental analysis indicates that ensemble models such as Random Forest generally perform better compared to individual classifiers. The model successfully identifies linguistic patterns and textual structures associated with fake news articles.

The results demonstrate that machine learning techniques can provide a reliable approach for detecting misinformation in online news content.

VIII. ADVANTAGES OF THE PROPOSED APPROACH

The proposed system offers several benefits:

- Automated detection of misleading news articles
- Reduction in manual verification effort
- Improved reliability of online information sources
- Potential integration with social media monitoring systems

IX. LIMITATIONS

Despite promising results, certain limitations remain:

- Fake news writing styles evolve continuously
- Model performance depends on dataset quality
- Some misleading articles may closely resemble genuine news

These challenges highlight the need for further research and improvement in detection techniques.

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 12, May 2026
Impact Factor: 8.2

X. FUTURE SCOPE

Future research can focus on integrating deep learning techniques such as neural networks and transformer-based language models. Combining textual analysis with image verification and user behaviour analysis could also improve detection accuracy.

Additionally, real-time fake news monitoring systems can be developed to analyse news content as it spreads across digital platforms.

XI. CONCLUSION

The increasing spread of misinformation on digital platforms highlights the importance of automated fake news detection systems. This study proposed a machine learning-based framework for identifying fake news through textual analysis.

By applying data preprocessing techniques and extracting meaningful features, machine learning models were trained to classify news articles accurately. The experimental results confirm that machine learning algorithms can effectively detect misleading information in online content.

The proposed system contributes to improving the credibility of digital information sources and provides a foundation for future research in misinformation detection.

REFERENCES

[1] K. Shu, A. Sliva, S. Wang, J. Tang, and H. Liu, “Fake News Detection on Social Media,” 2017.
[2] H. Ahmed, I. Traore, and S. Saad, “Detecting Opinion Spam and Fake News Using Text Classification,” 2018.
[3] X. Zhou and R. Zafarani, “A Survey of Fake News Detection Methods,” ACM Computing Surveys, 2020.
[4] C. Castillo, M. Mendoza, and B. Poblete, “Information Credibility on Twitter,” WWW Conference, 2011.
[5] S. Vosoughi, D. Roy, and S. Aral, “The Spread of True and False News Online,” Science Journal, 2018.