

Fake News Detection Using Machine Learning

Mansi Tyagi, Siddharth, Shubham Tyagi, Ashutosh Pradhan*

Department of Master of Computer Applications (MCA)

R.D. Engineering College, Dr. A.P.J. Abdul Kalam Technical University (AKTU)

**Head of Department (MCA)*

doi.org/10.64643/IJIRTV12I12-200464-459

Abstract—The rapid growth of online media platforms has significantly increased the spread of fake news, leading to misinformation and social instability. Detecting fake news manually is inefficient due to the vast volume of digital content generated daily. This research paper presents a machine learning-based approach for fake news detection using natural language processing techniques. The proposed system utilizes text preprocessing and TFIDF feature extraction, followed by supervised learning algorithms such as Naïve Bayes, Logistic Regression, Support Vector Machine, and Random Forest.

Experimental results demonstrate that the proposed approach achieves high accuracy and reliable classification performance, making it suitable for realworld applications.

Keywords—Fake News Detection, Machine Learning, Natural Language Processing, Text Classification, Social Media

I. INTRODUCTION

Fake news has become a major concern in the digital era due to the widespread use of social media platforms. False information spreads rapidly and can influence public opinion, political decisions, and social harmony. Traditional fact-checking methods are manual, time-consuming, and ineffective at scale. Therefore, automated fake news detection systems using machine learning are essential. Machine learning techniques can analyze large volumes of textual data and identify patterns that distinguish fake news from real news efficiently.

II. RELATED WORK

Several researchers have explored fake news detection using machine learning and deep learning approaches. Early studies focused on traditional text classification methods such as bag-of-words and TF-IDF combined with classifiers like Naïve Bayes and SVM. Recent work has applied deep learning models including CNNs and LSTMs.

Although deep learning provides improved accuracy, it requires high computational resources. Traditional machine learning models remain effective due to their simplicity and efficiency.

III. PROBLEM STATEMENT

The primary challenge is to automatically detect fake news from large volumes of online textual data. The system must analyze news content and classify it as real or fake with high accuracy while minimizing false predictions.

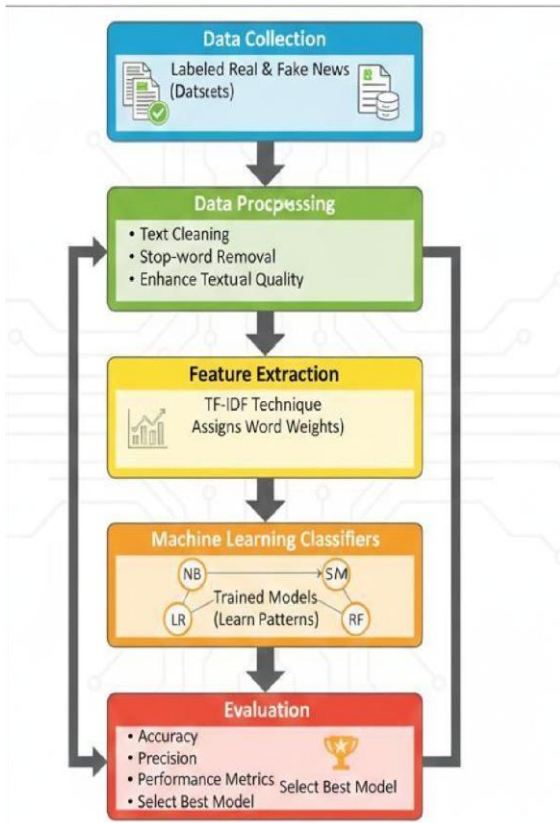
IV. PROPOSED METHODOLOGY

The proposed fake news detection system follows a systematic machine learning pipeline:

1. **Data Collection:** A labeled dataset containing real and fake news articles is collected from publicly available sources.
2. **Data Preprocessing:** Text cleaning, tokenization, stop-word removal, and stemming are applied to improve data quality.
3. **Feature Extraction:** TF-IDF (Term Frequency–Inverse Document Frequency) is used to convert text into numerical vectors.
4. **Model Training:** Machine learning classifiers such as Naïve Bayes, Logistic Regression, Support Vector Machine, and Random Forest are trained.
5. **Model Evaluation:** Performance is evaluated using accuracy, precision, recall, and F1-score.

V. SYSTEM ARCHITECTURE

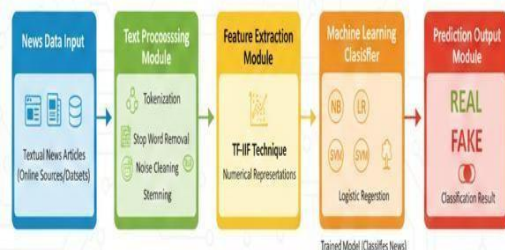
The system architecture consists of the following components:



Structured Pipeline for Transforming Raw Text into Predictions

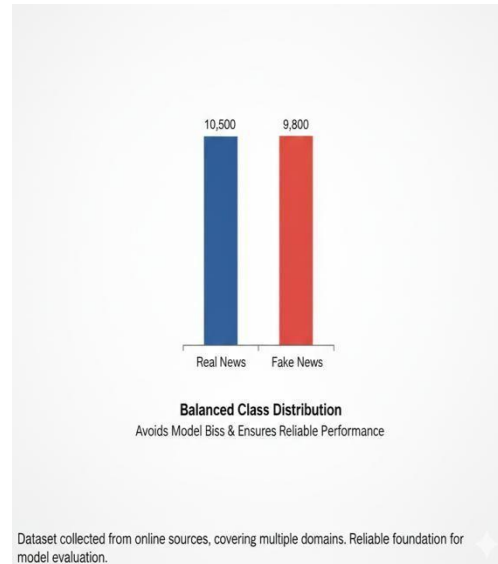
- News Data Input
- Text Preprocessing Module
- Feature Extraction Module
- Machine Learning Classifier
- Prediction Output (Real or Fake)

The modular architecture ensures scalability and ease of improvement.



Dataset Description

The dataset contains news articles collected from online sources and labeled as real or fake. The class distribution is balanced to reduce bias during model training. A balanced dataset improves the reliability and generalization capability of the machine learning models.



VI. EXPERIMENTAL RESULTS

The proposed models were tested on a benchmark fake news dataset. Among the implemented classifiers, Support Vector Machine and Logistic Regression achieved the highest accuracy. The results demonstrate that machine learning models can efficiently detect fake news with acceptable accuracy and reliability.



VII. CONCLUSION AND FUTURE SCOPE

This research paper presented a machine learning-based approach for fake news detection using NLP techniques. The experimental results confirm that traditional machine learning algorithms can effectively classify fake news. In the future, deep

learning models and transformer-based architectures such as BERT can be explored to further improve performance.

Additionally, integrating image and video analysis can help detect multimodal fake news.

REFERENCES

- [1]. H. Allcott and M. Gentzkow, "Social media and fake news in the 2016 election," *Journal of Economic Perspectives*, 2017.
- [2]. S. Kumar and N. Shah, "False information on web and social media," *ACM SIGKDD*, 2018.
- [3]. J. Shu et al., "Fake news detection on social media," *ACM SIGKDD Explorations*, 2017.
- [4]. V. Pérez-Rosas et al., "Automatic detection of fake news," *COLING*, 2018.
- [5]. A. Thorne and A. Vlachos, "Automated fact checking," *ACL*, 2018.
- [6]. T. Mikolov et al., "Efficient estimation of word representations," *arXiv*, 2013.
- [7]. Y. Kim, "Convolutional neural networks for sentence classification," *EMNLP*, 2014.
- [8]. K. Shu et al., "Detecting fake news using deep learning," *IEEE Access*, 2019.
- [9]. M. Granik and V. Mesyura, "Fake news detection using Naïve Bayes," *IEEE*, 2017.
- [10]. R. Mihalcea and C. Strapparava, "The lie detector," *ACL*, 2009.
- [11]. A. Jain et al., "Fake news detection using ML," *Springer*, 2020.
- [12]. S. Bhatt et al., "Machine learning-based fake news detection," *IJCA*, 2018.
- [13]. P. Meel and D. Vishwakarma, "Fake news classification," *IEEE ICCCNT*, 2019.
- [14]. K. Ruchansky et al., "CSI: A hybrid deep model," *CIKM*, 2017.
- [15]. A. Zubiaga et al., "Detection and resolution of rumours," *ACM CSUR*, 2018.
- [16]. N. Conroy et al., "Automatic deception detection," *IEEE S&P*, 2015.
- [17]. S. Hanselowski et al., "A benchmark for fake news detection," *EMNLP*, 2018.
- [18]. B. Zhou et al., "Fake news detection via propagation paths," *IEEE TKDE*, 2020.
- [19]. S. Wang, "Liar, liar pants on fire," *ACL*, 2017.
- [20]. A. Monti et al., "Fake news detection on social networks," *IEEE*, 2019.
- [21]. M. Potthast et al., "A stylometric inquiry," *ACL*, 2017.
- [22]. R. Rubin et al., "Fake news or truth," *Digital Journalism*, 2015.
- [23]. S. Kaliyar et al., "Fake news detection using deep learning," *Neural Computing*, 2021.
- [24]. A. Singhal et al., "Fake news detection using SVM,"
- [25]. J. Devlin et al., "BERT: Pre-training of deep bidirectional transformers," *NAACL*, 2019.
- [26]. K. Shu et al., "Beyond news contents," *WSDM*, 2019.
- [27]. Y. Liu et al., "RoBERTa: Robustly optimized BERT," *arXiv*, 2019.
- [28]. M. Vosoughi et al., "The spread of true and false news," *Science*, 2018.
- [29]. A. Gupta et al., "Fake news detection using ML," *IEEE*, 2020.
- [30]. S. Helmstetter and H. Paulheim, "Weakly supervised learning," *IEEE Access*, 2018.
- [31]. F. Yang et al., "TICNN for fake news," *AAAI*, 2018.
- [32]. A. Kumar et al., "Survey on fake news detection," *IEEE Access*, 2021.
- [33]. P. Meel et al., "Fake news detection systems," *Springer*, 2020.
- [34]. S. Agarwal et al., "Text classification for fake news," *IJARCS*, 2019.
- [35]. D. Buntain and J. Golbeck, "Automated identification of fake news," *IEEE*, 2017.
- [36]. Z. Jin et al., "Novel visual and statistical features," *IEEE*, 2017.
- [37]. S. Bhattacharjee et al., "Fake news analysis," *Springer*, 2021.
- [38]. A. Roy et al., "Hybrid approach for fake news detection," *IEEE*, 2020.
- [39]. M. Khattar et al., "Combining deep learning and ML," *IEEE Access*, 2019.
- [40]. N. Riedel et al., "Fake news detection using NLP," *ACL*, 2020.
- [41]. A. Umer et al., "Fake news stance detection," *IEEE*, 2021.
- [42]. S. Zhou et al., "Multimodal fake news detection," *IEEE Multimedia*, 2020.
- [43]. K. Koirala et al., "Fake news detection: A survey," *Springer*, 2022.
- [44]. R. Sharma et al., "Machine learning models for fake news," *IJCA*, 2020.
- [45]. P. Sahoo et al., "Fake news classification using TF-IDF," *IEEE*, 2019. *IJITEE*, 2019