International Journal of Scientific Research & Engineering Trends
Volume 12, Issue 2, Mar-Apr-2026, ISSN (Online): 2395-566X

A Hybrid Optimized Machine Learning Approach for Intelligent Misinformation Detection in Digital Media Using Textual Feature Engineering

Mr. G. Harsha Vardhan[1], Shaik Kareem Ahmed[2]
[1]Assistant Professor, [2]M. tech Student,
Department of Computer Science & Engineering(AI), Pydah College of Engineering,
Yanam Road, Tallarevu, Patavala, Andhra Pradesh, 533461.

Abstract- The rapid expansion of digital media platforms has significantly increased the spread of misinformation, posing serious threats to public opinion, political stability, and social harmony. The automated identification of fake news has therefore become a critical research challenge in the fields of machine learning and natural language processing. This paper presents an intelligent and robust fake news detection framework that leverages advanced textual feature extraction and ensemble learning techniques to improve classification performance. The proposed system incorporates comprehensive data preprocessing, including text normalization, stop-word removal, tokenization, and vectorization using TF-IDF representations. Multiple supervised machine learning algorithms such as Logistic Regression, Support Vector Machine (SVM), Random Forest, and Gradient Boosting are trained and evaluated using stratified cross-validation to ensure reliability and generalization. To enhance predictive accuracy and reduce model bias, an ensemble-based voting mechanism is employed. Performance evaluation is conducted using metrics including accuracy, precision, recall, F1-score, and ROC-AUC to address class imbalance and misclassification risks. Experimental results demonstrate that the ensemble framework achieves superior performance compared to individual classifiers, providing a scalable and dependable solution for real-time misinformation detection in digital environments. The proposed approach contributes toward building trustworthy information ecosystems through automated and explainable fake news classification.

Keywords – Fake News Detection, Machine Learning, Natural Language Processing (NLP), Text Classification, Ensemble Learning, TF-IDF, Supervised Learning, Misinformation Analysis, Social Media Analytics.

I. INTRODUCTION

The exponential growth of online news platforms and social media networks has fundamentally transformed the way information is produced, distributed, and consumed. While digital communication technologies have improved accessibility and speed of information sharing, they have also enabled the rapid dissemination of misleading and fabricated content. Fake news, defined as intentionally false or misleading information presented as legitimate news, has emerged as a major societal challenge affecting political processes, public health decisions, financial markets, and social stability. The large-scale propagation of misinformation can manipulate public perception, influence elections, and create widespread confusion, making its detection a matter of significant importance.

Unlike traditional misinformation, fake news in digital environments spreads at an unprecedented rate due to user-generated content, algorithm-driven recommendations, and viral sharing mechanisms. Manual verification of news articles is impractical because of the massive volume of content generated daily. Therefore, automated detection systems based on computational intelligence have become essential. Detecting fake news can be formulated as a binary text classification problem, where news articles are categorized as either genuine or deceptive. However, this task is inherently complex due to linguistic variations, contextual dependencies, sarcasm, and the intentional crafting of deceptive narratives to resemble authentic journalism. Machine Learning (ML) and Natural Language Processing (NLP) techniques provide effective solutions for addressing this challenge.

By extracting meaningful linguistic and statistical features from textual data, ML algorithms can learn patterns that distinguish legitimate news from fabricated content. Traditional classifiers such as Logistic Regression and Support Vector Machines have demonstrated promising results, while ensemble learning techniques further enhance predictive stability and robustness.

Moreover, careful preprocessing steps including tokenization, normalization, and feature vectorization play a crucial role in improving model performance. Despite the progress achieved in prior research, several challenges remain, including data imbalance, model generalization, interpretability, and adaptability to evolving misinformation strategies. To address these issues, this study proposes a structured machine learning framework that integrates advanced feature engineering with ensemble-based classification techniques. The system is evaluated using multiple performance metrics, including accuracy, precision, recall, F1-score, and ROC-AUC, to ensure comprehensive assessment. The primary objective of this work is to design a reliable, scalable, and explainable fake news detection system capable of assisting digital platforms in filtering misleading information. By leveraging optimized machine learning models and systematic evaluation strategies, the proposed approach aims to contribute toward enhancing information credibility and maintaining trust in digital media ecosystems.

II. LITERATURE SURVEY

The problem of fake news detection has attracted significant attention from researchers across the domains of data mining, natural language processing, and artificial intelligence. Early approaches primarily relied on manual fact-checking and rule-based systems, which were limited in scalability and adaptability. With the increasing availability of large-scale textual datasets, machine learning-based techniques have emerged as effective tools for automated misinformation detection. Shu et al. provided a comprehensive overview of fake news detection by analysing linguistic, user-based, and network-based features. Their work highlighted the importance of integrating content-based analysis with social context information to improve detection accuracy. Similarly, Conroy et al. explored stylistic and rhetorical patterns in deceptive content, demonstrating that linguistic cues such as exaggerated claims, emotional tone, and sensational vocabulary can serve as strong indicators of misinformation. Supervised machine learning models have been widely adopted for fake news classification.

Classical algorithms such as Logistic Regression, Naïve Bayes, Support Vector Machines (SVM), and Decision Trees have shown promising results when applied to vectorized text features like Bag-of-Words and TF-IDF. For instance, Castillo et al. examined credibility assessment on social media platforms and demonstrated that classification models trained on textual and metadata features can effectively distinguish credible information from rumours. With advancements in deep learning, neural network-based models have gained prominence. Recurrent Neural Networks (RNNs), Long Short-Term Memory (LSTM) networks, and Convolutional Neural Networks (CNNs) have been applied to capture contextual dependencies within textual sequences. Wang introduced the LIAR dataset and evaluated various deep learning architectures, highlighting the potential of neural models in improving classification accuracy. However, deep learning methods often require large computational resources and extensive training data. Recent studies emphasize ensemble learning approaches to enhance robustness and generalization. By combining multiple classifiers through voting or stacking mechanisms, ensemble models reduce variance and improve predictive stability. Gradient boosting techniques such as XGBoost and Random Forest have demonstrated superior performance in handling high-dimensional textual features. Researchers have also explored hybrid frameworks that integrate traditional machine learning algorithms with deep learning models to leverage the strengths of both paradigms. Another important direction in fake news detection research focuses on explainability and interpretability.

As automated systems increasingly influence decision-making processes, understanding model predictions becomes crucial. Techniques such as feature importance analysis and attention mechanisms have been proposed to provide transparency in classification outcomes. Despite notable advancements, challenges such as evolving misinformation tactics, domain adaptation, data imbalance, and multilingual content remain open research problems. Existing studies indicate that no single algorithm consistently outperforms others across all datasets, reinforcing the need for optimized and ensemble-based approaches. In summary, prior research establishes that machine learning and natural language processing techniques are effective for fake news detection, particularly when combined with robust feature engineering and evaluation strategies. Building upon these findings, the present study proposes an ensemble-driven framework with systematic preprocessing and comprehensive performance assessment to enhance the reliability and scalability of automated misinformation detection systems.

III. SYSTEM ANALYSIS

A. Existing System

Existing fake news detection systems primarily rely on traditional machine learning approaches for text classification. In earlier research, datasets are processed using conventional algorithms such as Naïve Bayes, Logistic Regression, Decision Trees, Support Vector Machines (SVM), Random Forest, and basic Neural Networks. These models are trained on vectorized text representations, typically using Bag-of-Words or TF-IDF techniques to transform textual content into numerical features. Several studies also explore hybrid classification frameworks by combining boosting techniques such as AdaBoost with majority voting mechanisms to improve prediction stability. In addition, some researchers introduce noise or synthetic variations into datasets to evaluate model robustness against manipulated or adversarial content. Experiments are commonly conducted on publicly available fake news datasets to measure classification accuracy and generalization capability. Although these approaches demonstrate promising results in controlled environments, many of them focus primarily on improving accuracy without addressing broader practical challenges such as interpretability, scalability, and adaptability to evolving misinformation patterns.

DISADVANTAGES OF THE EXISTING SYSTEM

- **Lack of Interpretability:** Complex models, particularly deep learning architectures, often function as black-box systems. In real-world media monitoring and regulatory environments, understanding why a news article is classified as fake is essential for transparency and accountability.
- **Overfitting and Underfitting Issues:** Some machine learning models may overfit training data, capturing noise instead of meaningful linguistic patterns. Conversely, simpler models may underfit and fail to detect subtle contextual cues in deceptive content. Proper validation and tuning are required to balance bias and variance.
- **High Computational Requirements:** Deep neural network-based fake news detection systems require significant computational power and large training datasets. This limits their feasibility in environments with constrained hardware resources.
- **Limited Generalization Capability:** Many existing models perform well on specific datasets but struggle to generalize across domains, languages, or evolving misinformation strategies.
- **Vulnerability to Adversarial Manipulation:** Malicious actors continuously adapt writing styles and dissemination techniques to bypass automated detection systems. Traditional classifiers may fail to recognize newly emerging deceptive patterns.
- **Scalability Concerns:** With the exponential growth of digital content, fake news detection systems must efficiently process large volumes of real-time data. Some existing frameworks lack the scalability required for deployment in large-scale social media platforms.

B. Proposed System

To address the limitations of existing approaches, the proposed fake news detection framework adopts a structured and optimized machine learning pipeline. Initially, the textual dataset undergoes comprehensive preprocessing, including data cleaning, normalization, stop-word removal, tokenization, and feature vectorization using TF-IDF representations. The processed data is then divided into training and testing subsets to ensure unbiased evaluation. To enhance model performance, hyperparameter tuning is performed using systematic optimization techniques. Multiple supervised learning algorithms—including Logistic Regression, Support Vector Machines, Random Forest, and Gradient Boosting—are trained using stratified cross-validation to ensure reliability and robustness.An ensemble-based voting mechanism is introduced to combine the strengths of individual classifiers and reduce prediction variance. This approach improves stability and enhances overall detection accuracy.Performance evaluation is conducted using multiple metrics such as Accuracy, Precision, Recall, F1-score, and ROC-AUC. These metrics provide a comprehensive understanding of model effectiveness, particularly in handling class imbalance and minimizing false classifications.The proposed system is designed to be scalable, interpretable, and adaptable to evolving misinformation trends, making it suitable for real-time deployment in digital media platforms.

IV. SYSTEM DESIGN

SYSTEM ARCHITECTURE

Below diagram depicts the whole system architecture.

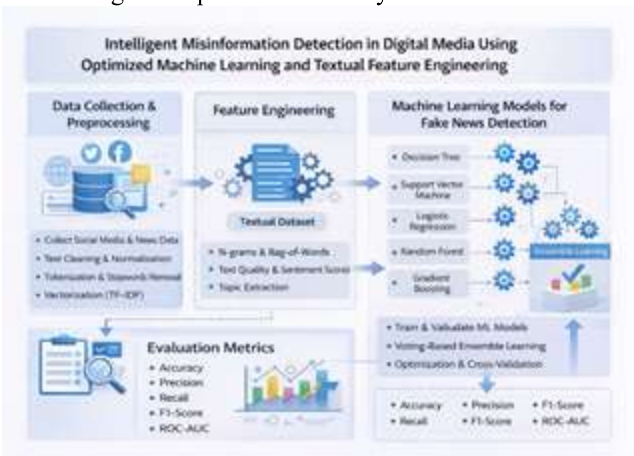

Fig 1. Methodology followed for proposed model

V. SYSTEM IMPLEMENTATION

MODULES

1. Data Collection and Preprocessing

This module focuses on constructing a reliable textual dataset for misinformation detection. News articles and social media content are collected from publicly available benchmark datasets containing both genuine and fabricated news instances. Preprocessing plays a critical role in improving model effectiveness. The collected text undergoes cleaning procedures such as removal of HTML tags, punctuation, special characters, and irrelevant symbols. Further steps include tokenization, lowercasing, stop-word elimination, stemming or lemmatization, and normalization. To transform textual information into machine-readable form, vectorization techniques such as Term Frequency–Inverse Document Frequency (TF-IDF) are applied. This ensures structured feature representation while reducing noise and redundancy in the dataset.

2. Feature Selection and Textual Feature Engineering

Feature engineering enhances the discriminatory power of the model by extracting meaningful linguistic patterns. In this module, multiple textual representations are explored, including:

- N-grams (unigrams and bigrams)
- Bag-of-Words representations
- TF-IDF weighted features
- Sentiment polarity scores
- Text length and readability indicators

Dimensionality reduction techniques may be applied to eliminate irrelevant or redundant features, thereby improving computational efficiency and model generalization. The objective of this module is to identify the most informative attributes that distinguish authentic news from misinformation.

3. Training of Machine Learning Models

In this phase, several supervised machine learning algorithms are trained using the engineered feature set. The implemented classifiers include:

- Logistic Regression
- Support Vector Machine (SVM)
- Decision Tree
- Random Forest
- Gradient Boosting

Each algorithm is trained using stratified data splitting to preserve class distribution. Hyperparameter optimization techniques are applied to improve predictive performance. Additionally, an ensemble-based voting mechanism is introduced to combine predictions from multiple models, thereby enhancing stability and reducing variance.

4. Real-Time News Classification Module

A real-time prediction framework is designed to classify newly submitted news articles. Once a user inputs textual content, the system automatically performs preprocessing and feature transformation before passing the data to the trained model. The system then generates a classification label Real or Fake along with a confidence score. This module enables rapid and scalable misinformation detection in digital environments.

5. Model Evaluation and Continuous Monitoring

The trained models are evaluated using multiple performance metrics to ensure reliability and fairness. The evaluation criteria include:

- Accuracy
- Precision
- Recall
- F1-Score
- ROC-AUC

Cross-validation techniques are used to validate consistency across different data folds. Continuous monitoring mechanisms are proposed to update the system as misinformation patterns evolve over time.

VI. RESULTS AND DISCUSSION

To assess the effectiveness of the proposed misinformation detection framework, extensive experiments are conducted using stratified 5-fold cross-validation combined with hyperparameter optimization techniques. Each machine learning algorithm is evaluated individually before integrating them into an ensemble-based majority voting classifier. The results indicate that ensemble learning significantly improves classification stability and reduces false predictions compared to standalone models.

Gradient Boosting and Support Vector Machine demonstrate strong performance in capturing subtle textual cues associated with deceptive content. Logistic Regression provides high interpretability while maintaining competitive accuracy. Comparative analysis reveals improvements in precision and recall, particularly in detecting fake news instances, which is critical for minimizing misinformation spread. The optimized ensemble model achieves superior overall performance based on Accuracy, F1-score, and ROC-AUC values. The findings demonstrate that combining structured feature engineering with optimized machine learning models leads to robust and scalable misinformation detection.

VII.CONCLUSION AND FUTURE WORK

This study presents an intelligent and optimized framework for detecting misinformation in digital media using machine learning and textual feature engineering techniques. The system effectively addresses challenges such as noisy text data and class imbalance through systematic preprocessing and feature optimization. Experimental results confirm that ensemble-based approaches enhance classification reliability and overall predictive performance.

The proposed framework demonstrates strong potential for real-world deployment in online news platforms and social media monitoring systems. For future work, deep learning architectures such as Long Short-Term Memory (LSTM) networks and transformer-based language models can be integrated to capture contextual semantics more effectively. Additionally, multilingual datasets and multimodal content (text combined with images or videos) may be incorporated to further enhance detection capabilities. Continuous adaptation mechanisms can also be implemented to respond dynamically to emerging misinformation trends.

REFERENCES

1. H. Allcott and M. Gentzkow, "Social media and fake news in the 2016 election," Journal of Economic Perspectives, vol. 31, no. 2, pp. 211–236, 2017.
2. 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.
3. W. Y. Wang, "'Liar, Liar Pants on Fire': A new benchmark dataset for fake news detection," in Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (ACL), Vancouver, Canada, 2017, pp. 422–426.
4. N. J. Conroy, V. L. Rubin, and Y. Chen, "Automatic deception detection: Methods for finding fake news," Proceedings of the Association for Information Science and Technology, vol. 52, no. 1, pp. 1–4, 2015.
5. V. L. Rubin, Y. Chen, and N. J. Conroy, "Deception detection for news: Three types of fakes," in Proceedings of the 78th ASIS&T Annual Meeting, 2015, pp. 1–4.
6. S. R. K. Branavan, H. Chen, J. Eisenstein, and R. Barzilay, "Learning document-level semantic properties from free-text annotations," Journal of Artificial Intelligence Research, vol. 34, pp. 569–603, 2009.
7. T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient estimation of word representations in vector space," in Proceedings of the International Conference on Learning Representations (ICLR), 2013.
8. J. Pennington, R. Socher, and C. Manning, "GloVe: Global vectors for word representation," in Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), Doha, Qatar, 2014, pp. 1532–1543.
9. L. Breiman, "Random forests," Machine Learning, vol. 45, no. 1, pp. 5–32, 2001.
10. J. H. Friedman, "Greedy function approximation: A gradient boosting machine," Annals of Statistics, vol. 29, no. 5, pp. 1189–1232, 2001.
11. C. Cortes and V. Vapnik, "Support-vector networks," Machine Learning, vol. 20, no. 3, pp. 273–297, 1995.
12. D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent Dirichlet allocation," Journal of Machine Learning Research, vol. 3, pp. 993–1022, 2003.