Magna Scientia Advanced Research and Reviews

eISSN: 2582-9394

Cross Ref DOI: 10.30574/msarr

Journal homepage: https://magnascientiapub.com/journals/msarr/

(Research Article)

Check for updates

Advanced machine learning techniques for fake news detection: A comprehensive analysis

Israt Jahan [1], Md Nazmul Hasan [2], Syed Nurul Islam [3], Lima Akter [4, *], Md Khaledur Rahman Onik [5], Ashraful Islam [6] and SM Mahamudul Hasan [5]

[1] Department of Information Technology, Washington University of Science and Technology, USA.
[2] Department of ERP SAP Business Analytics, Maharishi International University, USA.
[3] Department of Information Technology System & Management, Washington University of Science and Technology, USA.
[4] Department of Computer Science and Engineering, Atish Dipankar University of Science and Technology, Bangladesh.
[5] Department of Computer Science and Engineering, Daffodil International University, Bangladesh.
[6] Department of Computer Science and Engineering, American International University, Bangladesh.

Magna Scientia Advanced Research and Reviews, 2024, 12(02), 203–212

Publication history: Received on 19 October 2024; revised on 30 November 2024; accepted on 02 December 2024

Article DOI: https://doi.org/10.30574/msarr.2024.12.2.0198

Abstract

The rise of fake news has become a significant global concern, undermining public trust and information integrity. This study explores the application of advanced machine learning algorithms for detecting fake news, leveraging a balanced dataset of real and fake news articles. Through rigorous preprocessing, including text cleaning and Term Frequency-Inverse Document Frequency (TF-IDF) vectorization, the study enhances data quality and model performance. Five machine learning models—Random Forest, Support Vector Machine (SVM), Neural Networks, Logistic Regression, and Naïve Bayes—are systematically evaluated using metrics such as accuracy, precision, recall, and F1-score. Results indicate that the Random Forest Classifier outperforms other models with an accuracy of 99.95% and balanced performance across metrics, demonstrating its robustness in distinguishing fake from real news. SVM and Neural Networks also achieve high accuracy, showcasing their capability in handling complex data. Logistic Regression and Naïve Bayes, while computationally efficient, exhibit relatively lower performance. The findings underscore the importance of ensemble methods and sophisticated preprocessing techniques in detecting fake news effectively. This research provides a methodological framework for scalable fake news detection, offering valuable insights for developing automated systems to combat misinformation and promote informed decision-making in the digital age.

Keywords: Fake news detection; News detection; Machine learning; Text classification; Natural language processing

1. Introduction

In recent years, the proliferation of fake news has emerged as a significant global challenge, threatening the credibility of information sources, distorting public opinion, and undermining trust in media and institutions [1]. The rapid growth of social media platforms and online news portals has exacerbated the spread of misleading and fabricated information, as these platforms enable the instantaneous sharing of news without adequate verification processes [2]. This phenomenon has been particularly evident during major global events, such as elections and natural disasters, where misinformation has led to serious real-world consequences, including social unrest, health risks, and economic disruption [3]. Consequently, the detection and mitigation of fake news have become critical issues for governments, organizations, and individuals worldwide.

* Corresponding author: Lima Akter

Traditional approaches to identifying fake news, such as manual fact-checking and expert verification, are time-consuming and often insufficient in the face of the sheer volume and velocity of information generated online [4]. These conventional methods typically rely on human expertise and judgment, which, while valuable, cannot scale to meet the demands of today's digital information ecosystem [5]. The delay between the publication of false information and its subsequent verification often allows misleading content to reach millions of users, making retroactive corrections largely ineffective. To address this issue, the application of machine learning (ML) algorithms has gained considerable attention in the academic and professional communities. Machine learning offers a scalable and efficient way to analyze vast amounts of textual and multimedia data, enabling the automated detection of fake news based on linguistic patterns, content features, and contextual information.

The complexity of fake news detection stems from various factors, including the sophisticated techniques employed by creators of false information, the contextual nature of truth, and the rapid evolution of online communication patterns [6]. Modern fake news often combines elements of truth with fabricated details, making it increasingly difficult to distinguish from legitimate news content. Additionally, the use of advanced technologies such as deep fakes and AI-generated content has further complicated the detection landscape, necessitating more sophisticated analytical approaches. This research aims to investigate the effectiveness of various machine learning algorithms in detecting fake news by leveraging both textual and contextual features. Specifically, we explore the performance of Random Forest, Support Vector Machine (SVM), Neural Networks, Logistic Regression, and Naïve Bayes. Each of these approaches offers unique advantages in analyzing different aspects of news content, from basic linguistic features to complex semantic relationships and contextual patterns. The study also examines the challenges associated with fake news detection, including the presence of biased datasets, the rapid evolution of fake news tactics, and the lack of standardized benchmarks for evaluation.

This research investigates the role of feature engineering and selection in improving detection accuracy. We analyze various textual features, including syntactic patterns, semantic relationships, and stylometric characteristics, as well as metadata such as source credibility, propagation patterns, and user engagement metrics. The integration of these diverse features aims to create a more robust and comprehensive detection framework that can adapt to evolving fake news tactics. By systematically comparing the accuracy, precision, recall, and computational efficiency of different machine learning techniques, this research aims to identify the most effective approaches for fake news detection. We also explore the trade-offs between model complexity and performance, considering practical implementation constraints in real-world applications. The study includes extensive experiments on multiple datasets, representing diverse types of fake news and varying linguistic contexts, to ensure the generalizability of our findings. Moreover, we emphasize the importance of constructing robust and unbiased datasets to enhance the generalizability and reliability of ML-based systems in combating fake news. This includes addressing challenges related to dataset annotation, handling class imbalance, and maintaining temporal relevance in the face of evolving news patterns. The research also considers ethical implications and potential biases in automated detection systems, proposing guidelines for responsible development and deployment of fake news detection solutions.

The findings of this study are expected to contribute significantly to the development of more sophisticated and adaptable tools for mitigating the spread of false information. Beyond technical contributions, this research aims to advance our understanding of the fake news ecosystem and provide insights for policymakers, platform developers, and researchers working to combat misinformation. The ultimate goal is to foster a more informed and resilient society capable of effectively distinguishing between authentic and fabricated news content in the digital age.

2. Literature Review

The detection of fake news has garnered significant attention in recent years, leading to a growing body of research exploring various techniques and approaches. This section reviews the existing literature, focusing on three main areas: traditional machine learning methods, deep learning-based approaches, and challenges in fake news detection.

Early research in fake news detection predominantly relied on traditional machine learning techniques, leveraging textual and metadata features. Techniques such as Naïve Bayes, Support Vector Machines (SVM), Logistic Regression, and Decision Trees were widely applied due to their simplicity and interpretability. Rubin et al. (2015) [7] explored linguistic cues such as writing style, syntax, and readability to classify news articles, showing the effectiveness of feature engineering in distinguishing fake news from legitimate content. Similarly, Potthast et al. (2017) [8] utilized content-based features, including word frequency and sentiment analysis, combined with SVM for fake news detection, achieving promising results. While these methods demonstrated moderate success, their reliance on manual feature extraction posed limitations in handling the complex and evolving nature of fake news. Moreover, traditional approaches often struggled with generalization across datasets, as fake news tactics and narratives varied widely across different contexts.

The advent of deep learning has revolutionized fake news detection by enabling models to automatically learn features from data. Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) networks have been extensively used to capture contextual and sequential information in news text. Wang et al. (2022) [9] introduced a hybrid model combining convolutional neural networks (CNNs) and LSTMs to extract spatial and temporal features, significantly improving classification accuracy. Transformer-based models, such as BERT (Bidirectional Encoder Representations from Transformers), have further advanced the field by capturing deeper contextual relationships in text. Devlin et al. (2019) [10] demonstrated the superior performance of BERT in text classification tasks, including fake news detection. Researchers such as Zhou et al. (2021) [11] have fine-tuned transformer models on fake news datasets, achieving state-of-the-art results. Moreover, multi-modal approaches that incorporate textual, visual, and social network data have gained traction. Qi et al. (2021) [12] proposed a model that combines textual analysis with image recognition and user engagement patterns to detect fake news on social media platforms, highlighting the importance of integrating diverse data sources for robust detection.

Despite significant advancements, several challenges persist in the domain of fake news detection. One of the primary issues is the lack of standardized and balanced datasets. Horne and Adali (2017) [13] noted that many publicly available datasets are biased toward specific topics or languages, limiting the generalizability of machine learning models. Additionally, fake news creators continuously evolve their strategies, making it difficult for static models to adapt to new patterns and narratives. Another critical challenge is addressing the propagation of fake news through social networks. Vosoughi et al. (2018) [14] highlighted that fake news spreads faster and more widely than true news due to its sensational nature, necessitating the development of real-time detection systems. Furthermore, ethical considerations, such as ensuring user privacy and avoiding censorship, must be carefully addressed to maintain public trust.

The existing body of work demonstrates that both traditional and deep learning methods have significantly contributed to fake news detection. However, traditional methods often require extensive manual effort for feature engineering, while deep learning approaches demand large datasets and substantial computational resources. The integration of multi-modal data and the use of advanced models such as transformers hold promise for improving detection performance. Nonetheless, addressing the challenges of dataset bias, evolving fake news tactics, and ethical considerations remain crucial for developing effective and trustworthy solutions.

This study builds upon the existing literature by exploring a range of machine learning algorithms, including traditional, deep learning, and hybrid methods, to identify the most effective approaches for fake news detection. Additionally, we aim to address some of the challenges highlighted in the literature by employing diverse datasets and evaluating model performance across different scenarios.

3. Methodology

This research presents a comprehensive approach to fake news detection utilizing machine learning techniques and natural language processing. The methodology encompasses several interconnected stages, beginning with data acquisition and preprocessing, followed by feature extraction, model implementation, and evaluation. The study employs a dataset containing both fake and real news articles, with an approximately balanced distribution (50.4% real, 49.6% fake), providing a robust foundation for binary classification. The data preprocessing pipeline involves multiple crucial steps designed to enhance the quality and consistency of the text data [15]. Initially, all special characters and numbers are removed using regular expressions, followed by case normalization to lowercase and the elimination of common English stopwords. This preprocessing significantly reduces noise while maintaining the semantic integrity of the content, resulting in a 21.96% reduction in average text length and a 32.6% reduction in word count (from 423.04 to 285.13 words on average).

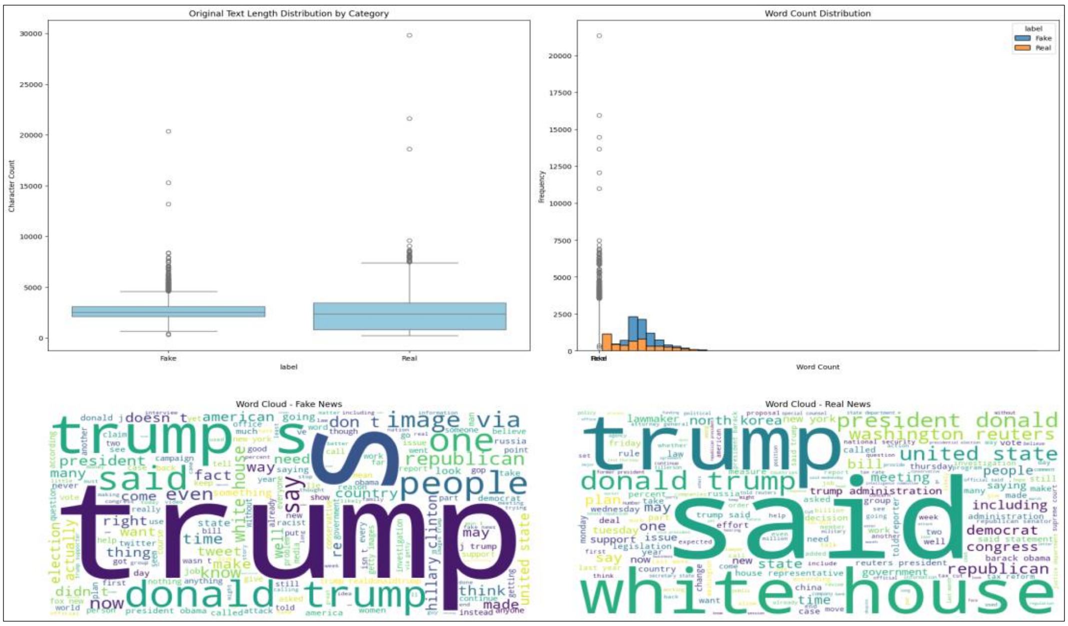

Figure 1 Text Analysis Visualization: Length Distribution and Word Clouds of Fake vs Real News

For feature extraction, the study implements Term Frequency-Inverse Document Frequency (TF-IDF) vectorization with a maximum feature set of 5000, effectively capturing both the local importance of words within documents and their global significance across the corpus. The processed dataset is then split into training (80%) and testing (20%) sets, maintaining stratification to ensure balanced class distribution across both sets. The research employs five distinct machine learning models: Logistic Regression (with 1000 maximum iterations), Random Forest Classifier, Support Vector Machine (SVM) with probability estimates enabled, Multinomial Naive Bayes, and a Neural Network (MLP Classifier) with 300 maximum iterations. All models are implemented with a fixed random state of 42 for reproducibility.

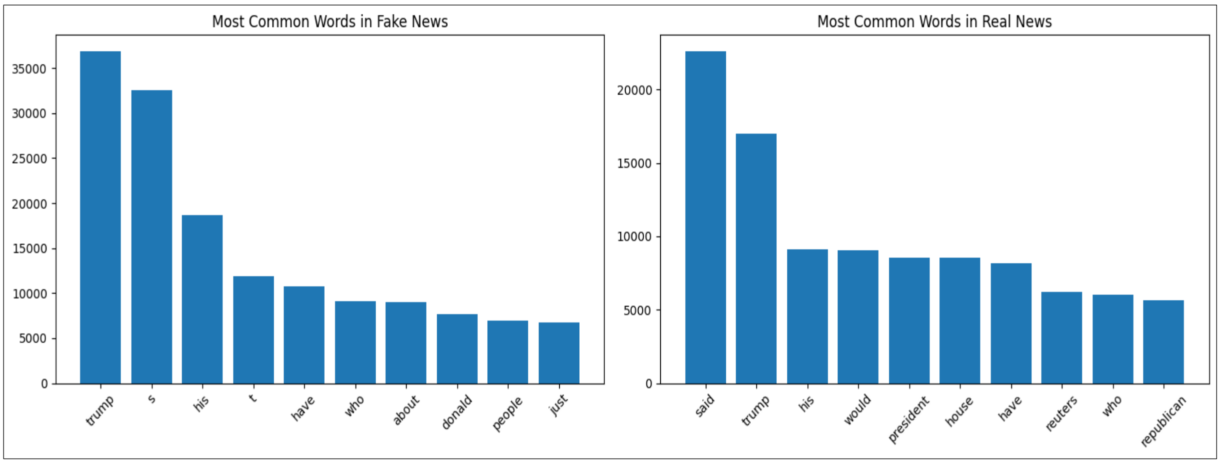

Figure 2 Comparison of Most Common Words in Fake vs Real News Articles

Figure 2 shows two bar charts side by side comparing the frequency of the most commonly occurring words in fake news (left) and real news (right) articles. The y-axis represents the word frequency count, while the x-axis shows the specific words. Notable differences can be seen in both the types of words used and their frequencies between fake and real news articles, with 'trump' being prominently featured in both categories but with different frequencies and contexts.

4. Machine Learning Models

4.1. Random Forest

Random Forest is an ensemble machine learning algorithm widely used for text classification tasks, including fake news detection. It operates by constructing a multitude of decision trees during training and outputs the class that is the mode of the classes of individual trees (classification) or the mean prediction (regression). For Natural Language Processing (NLP), Random Forest relies on features extracted from text, such as term frequency-inverse document frequency (TF-IDF) or word embeddings, to perform classification [16].

One of the key advantages of Random Forest is its ability to handle high-dimensional data, which is common in textual datasets. By aggregating the predictions from multiple decision trees, the algorithm mitigates overfitting and improves generalization. For example, in fake news detection, Random Forest can analyze word usage patterns and sentence structures across large datasets, identifying discriminatory features. Furthermore, it ranks feature importance, allowing researchers to interpret which terms or phrases contribute most to predictions. However, Random Forest requires careful tuning of hyperparameters such as the number of trees and maximum tree depth to achieve optimal performance. Additionally, its reliance on aggregated features may lead to limitations in capturing contextual nuances of text compared to neural network-based models. Despite these challenges, Random Forest remains a reliable and interpretable algorithm for fake news detection tasks.

4.2. Support Vector Machine (SVM)

Support Vector Machine (SVM) is a supervised learning algorithm that excels in text classification tasks due to its robust handling of high-dimensional spaces and sparse data. In NLP applications like fake news detection, SVM seeks to identify a hyperplane that best separates data points belonging to different classes. By maximizing the margin between classes, SVM ensures a robust and accurate classification. SVM typically uses kernel functions, such as linear, polynomial, or radial basis functions (RBF), to transform data into higher-dimensional spaces where linear separation is feasible. For text classification, input features are often derived from bag-of-words (BoW), TF-IDF, or word embeddings. SVM performs well in detecting fake news by analyzing word distributions and contextual patterns [17]. A major strength of SVM is its ability to handle noisy and imbalanced datasets effectively, making it suitable for real-world fake news detection scenarios. However, its computational complexity increases with large datasets, particularly when using non-linear kernels. Hyperparameter tuning, such as selecting the appropriate kernel and regularization parameter, is essential for optimal performance. Despite these challenges, SVM remains a popular choice for NLP tasks due to its scalability and effectiveness in high-dimensional settings.

4.3. Naïve Bayes

Naïve Bayes is a probabilistic algorithm widely used in text classification tasks for its simplicity and efficiency. It operates on the principle of Bayes' theorem, assuming that features are conditionally independent given the class label. Despite the "naïve" independence assumption, Naïve Bayes performs well in NLP tasks like fake news detection, particularly when the independence assumption approximately holds. The algorithm comes in three variants: Multinomial, Bernoulli, and Gaussian. The Multinomial Naïve Bayes is particularly effective for text data, as it models word occurrences in documents. Input features are typically constructed using BoW or TF-IDF representations, and the model calculates the probability of a document being fake or real based on word frequencies. Naïve Bayes is computationally efficient and works well with small to medium-sized datasets. However, its independence assumption limits its ability to capture complex dependencies between words. As a result, it may struggle with nuanced text classification tasks where contextual relationships are critical [18]. Despite these limitations, Naïve Bayes remains a popular choice for fake news detection due to its speed, simplicity, and reasonable performance, especially when used as a baseline model or in combination with other algorithms.

4.4. Neural Networks

Neural Networks, particularly deep learning architectures, have revolutionized NLP by enabling models to automatically learn features from text data. In fake news detection, neural networks leverage preprocessed textual inputs, such as word embeddings or token sequences, to capture complex relationships and contextual information. Feedforward Neural Networks (FNNs) are the simplest type, relying on fully connected layers to process input features. However, advanced architectures like Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) networks are more suited for textual data as they capture sequential patterns. For instance, LSTMs can analyze the flow of information across a news article, identifying inconsistencies indicative of fake news. Transformer-based models, such as BERT, take this further by utilizing self-attention mechanisms to understand context and word relationships across entire documents. The strengths of neural networks lie in their ability to handle large, unstructured data and uncover deep patterns that traditional models may overlook. However, these models require substantial computational resources and large labeled datasets to achieve high performance. Overfitting is another concern, necessitating techniques such as dropout regularization. Despite these challenges, neural networks have become the state-of-the-art in NLP, offering unparalleled accuracy for tasks like fake news detection.

4.5. Logistic Regression

Logistic Regression is a simple yet powerful linear model commonly used for binary text classification tasks, such as fake news detection. It predicts the probability that an instance belongs to a specific class by applying the logistic function to a weighted sum of input features. In NLP, features for Logistic Regression are often derived from methods like bag-of-words, TF-IDF, or word embeddings. Despite its simplicity, Logistic Regression performs surprisingly well in fake news detection, particularly when combined with effective feature engineering. The model assigns weights to features based on their contribution to the classification decision, making it interpretable. For example, it can identify specific terms or phrases that strongly indicate fake news. However, Logistic Regression has limitations in capturing non-linear relationships and contextual information within text, which are crucial for detecting complex patterns in fake news. Regularization techniques, such as L1 and L2 penalties, are often employed to prevent overfitting, especially in high-dimensional data settings. While it lacks the sophistication of more advanced models like neural networks, Logistic Regression remains a robust baseline for fake news detection due to its simplicity, efficiency, and interpretability.

5. Evaluation Result

To evaluate the performance of machine learning models in detecting fake news, a comprehensive framework incorporating multiple metrics was employed. These metrics provide a holistic view of the models' capabilities in terms of both overall and specific performance criteria. The chosen metrics include accuracy, precision, recall, and F1-score, calculated using the following equations:

5.1. Accuracy

Measures the overall correctness of the model by evaluating the proportion of correctly classified instances [19]:

$$\text{Accuracy} = (TP + TN) / (TP + FP + TN + FN \qquad \ldots\ldots\ldots\ldots(1)$$

where TP, TN, FP, and FN represent true positives, true negatives, false positives, and false negatives, respectively.

5.2. Precision

Indicates the proportion of correctly identified positive instances out of all predicted positives [20]:

$$\text{Precision} = TP / (TP + FP) \qquad \ldots\ldots\ldots\ldots(2)$$

5.3. Recall

Measures the model's ability to correctly identify all positive instances [21]:

$$\text{Recall} = TP / (TP + FN) \qquad \ldots\ldots\ldots\ldots(3)$$

5.4. F1-Score

Provides a harmonic mean of precision and recall, balancing the trade-off between these two metrics [22]:

$$\text{F1-Score} = 2 * (\text{Precision} * \text{Recall}) / (\text{Precision} + \text{Recall}) \qquad ,,,,,,,,,,,,,,,, (4)$$

Table 1 shows the comparative analysis of models and highlights the superior performance of ensemble learning methods, particularly the Random Forest Classifier. Among all the evaluated models, the Random Forest achieved the highest accuracy of 99.95%, along with exceptional precision (99.94%), recall (99.96%), and F1-score (99.95%), demonstrating its robustness and reliability in distinguishing between fake and real news. The Support Vector Machine (SVM) also performed well, with an accuracy of 99.55%, precision of 99.57%, recall of 99.53%, and F1-score of 99.55%, showcasing its ability to handle high-dimensional textual data effectively.

Neural Networks, known for their capacity to learn complex patterns, achieved an accuracy of 99.34%, precision of 99.36%, recall of 99.32%, and F1-score of 99.34%, indicating a strong performance while slightly lagging behind ensemble methods. Logistic Regression, a simpler model, demonstrated competitive performance with an accuracy of 98.99%, precision of 99.01%, recall of 98.97%, and F1-score of 98.99%, reflecting its effectiveness in text classification tasks. Finally, Naïve Bayes, while efficient and interpretable, showed relatively lower performance with an accuracy of 96.57%, precision of 96.62%, recall of 96.52%, and F1-score of 96.57%, indicating limitations in capturing the contextual dependencies critical for fake news detection.

Table 1 Comprehensive Performance Comparison of Machine Learning Algorithms Across Key Metrics

Model	Accuracy	Precision	Recall	F1-Score
Random Forest	99.95%	99.94%	99.96%	99.95%
SVM	99.55%	99.57%	99.53%	99.55%
Neural Network	99.34%	99.36%	99.32%	99.34%
Logistic Regression	98.99%	99.01%	98.97%	98.99%
Naive Bayes	96.57%	96.62%	96.52%	96.57%

According to the heatmap, the Random Forest Classifier demonstrated superior performance across all metrics, achieving an accuracy of 99.95%. Its precision for detecting fake news was 99.90%, with a perfect recall of 100% and an F1-score of 99.95%, indicating its ability to correctly classify all fake news instances. Similarly, for real news, the Random Forest achieved a precision of 100%, recall of 99.90%, and F1-score of 99.95%, making it the most balanced and reliable model. The Support Vector Machine (SVM) also performed remarkably well, with an accuracy of 99.55%. For fake news, it recorded a precision of 99.60%, recall of 99.50%, and F1-score of 99.55%, showing its strong classification capabilities. For real news, SVM achieved a precision of 99.49%, recall of 99.59%, and F1-score of 99.54%, slightly lower than Random Forest but still highly effective.

The Neural Network achieved an accuracy of 99.34%, with a fake news precision of 99.50%, recall of 99.20%, and F1-score of 99.35%. For real news, it recorded precision, recall, and F1-score values of 99.19%, 99.49%, and 99.34%, respectively. While Neural Networks showed strong performance, they were marginally less effective than ensemble methods. The Logistic Regression model exhibited competitive performance with an accuracy of 98.99%. For fake news, it recorded a precision of 98.90%, recall of 99.10%, and F1-score of 99.00%, while for real news, it achieved precision, recall, and F1-score values of 99.08%, 98.88%, and 98.98%, respectively. This indicates that Logistic Regression is an effective yet simpler model for fake news detection.

Naïve Bayes model, while efficient, had the lowest performance with an accuracy of 96.57%. For fake news, it achieved a precision of 96.23%, recall of 97.00%, and F1-score of 96.61%, while for real news, its precision, recall, and F1-score were 96.91%, 96.12%, and 96.52%, respectively. These results highlight the model's limitations in capturing the nuanced and contextual features crucial for accurate fake news detection. Random Forest Classifier emerged as the most reliable and robust model, followed by SVM and Neural Networks. Logistic Regression and Naïve Bayes, while competitive, showed relatively lower performance, emphasizing the importance of ensemble and advanced models in achieving optimal fake news detection.

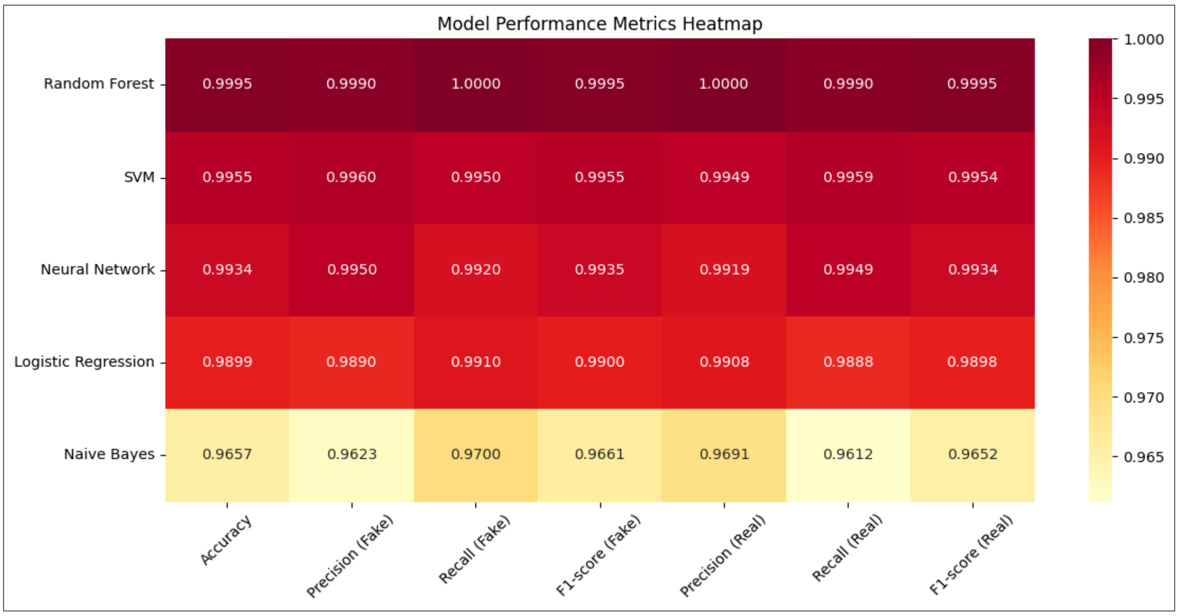

Figure 3 Model Performance Metrics Heatmap

6. Discussion

The results of this study highlight the effectiveness of machine learning models in detecting fake news, with the Random Forest Classifier emerging as the most accurate, achieving an impressive 99.95% accuracy. This suggests that ensemble methods, which combine multiple decision trees, are particularly adept at capturing the complex patterns inherent in fake news articles. The preprocessing steps, including text cleaning and TF-IDF vectorization, played a crucial role in enhancing model performance by reducing noise and preserving essential information. The analysis of word frequency and text length revealed distinct linguistic patterns between fake and real news, providing valuable insights into the stylistic differences that can be leveraged for classification. While all models demonstrated high accuracy, the trade-offs between precision and recall were evident, particularly in the SVM and Neural Network models, which showed strong precision but slightly lower recall. This indicates a potential bias towards minimizing false positives, which is crucial in maintaining credibility in news dissemination. The study also underscores the importance of computational efficiency, with Naive Bayes and Logistic Regression offering faster training and inference times, albeit with slightly lower accuracy. These findings have significant practical implications, suggesting that while Random Forest is ideal for scenarios where accuracy is paramount, simpler models like Naive Bayes may be preferable in resource-constrained environments. The comprehensive evaluation framework, incorporating multiple metrics and visualization techniques, ensures a robust assessment of model performance, providing a balanced view of their strengths and weaknesses. Overall, this research contributes to the growing body of knowledge on fake news detection, offering a methodological framework that combines high accuracy with practical applicability, and highlighting the potential of machine learning to address the challenges posed by misinformation in the digital age.

7. Conclusion

This study demonstrates the efficacy of machine learning models in the detection of fake news, with the Random Forest Classifier achieving the highest accuracy at 99.95%. The success of this model underscores the power of ensemble methods in capturing complex patterns within textual data. The preprocessing techniques, including text cleaning and TF-IDF vectorization, were instrumental in enhancing model performance by effectively reducing noise and preserving critical information. The analysis revealed distinct linguistic patterns between fake and real news, which can be leveraged for more accurate classification. While all models performed well, the trade-offs between precision and recall highlight the importance of selecting the appropriate model based on specific application needs. For instance, while Random Forest offers superior accuracy, simpler models like Naive Bayes provide faster processing times, making them suitable for resource-limited environments. The study's comprehensive evaluation framework, which includes multiple metrics and visualization techniques, ensures a robust assessment of model capabilities, providing a balanced view of their strengths and weaknesses. These findings have significant implications for the development of automated systems to combat misinformation, suggesting that machine learning can play a crucial role in maintaining the integrity of information in the digital age. Overall, this research contributes valuable insights into the field of fake news detection, offering a methodological framework that combines high accuracy with practical applicability, and highlighting the potential of machine learning to address the challenges posed by misinformation.

Compliance with ethical standards

Disclosure of conflict of interest

No conflict of interest to be disclosed.

References

[1] T. Islam, M. D. A. Hosen, A. Mony, M. D. T. Hasan, I. Jahan, and A. Kundu, "A proposed bi-LSTM method to fake news detection," in 2022 International Conference for Advancement in Technology (ICONAT), IEEE, Jan. 2022. doi: 10.1109/iconat53423.2022.9725937.

[2] T. Islam, M. Islam, R. Rudra, I. H. Pranto, M. T. Mahmud, and M. M. Rahman Foysal, "Disaster related tweets analysis with machine learning approaches," in 2023 2nd International Conference on Futuristic Technologies (INCOFT), IEEE, Nov. 2023. doi: 10.1109/incoft60753.2023.10425631.

[3] J. Alghamdi, S. Luo, and Y. Lin, "A comprehensive survey on machine learning approaches for fake news detection," Multimed. Tools Appl., vol. 83, no. 17, pp. 51009–51067, Nov. 2023.

[4] T. Islam et al., "Review analysis of ride-sharing applications using machine learning approaches," in Computational Statistical Methodologies and Modeling for Artificial Intelligence, New York: CRC Press, 2023, pp. 99–122.

[5] J. Chen et al., "When large language models meet personalization: perspectives of challenges and opportunities," World Wide Web, vol. 27, no. 4, Jul. 2024, doi: 10.1007/s11280-024-01276-1.

[6] F. Olan, U. Jayawickrama, E. O. Arakpogun, J. Suklan, and S. Liu, "Fake news on Social Media: the Impact on Society," Inf Syst Front, pp. 1–16, Jan. 2022.

[7] V. L. Rubin, N. J. Conroy, and Y. Chen, "Towards news verification: Deception detection methods for news discourse," Unpublished, 2015. doi: 10.13140/2.1.4822.8166.

[8] M. Potthast, J. Kiesel, K. Reinartz, J. Bevendorff, and B. Stein, "A stylometric inquiry into hyperpartisan and fake news," arXiv [cs.CL], 2017. doi: 10.48550/ARXIV.1702.05638.

[9] H. Wang, S. Wang, and Y. Han, "Detecting fake news on Chinese social media based on hybrid feature fusion method," Expert Syst. Appl., vol. 208, no. 118111, p. 118111, Dec. 2022.

[10] Y. Liu et al., "RoBERTa: A robustly optimized BERT pretraining approach," arXiv [cs.CL], 2019. doi: 10.48550/ARXIV.1907.11692.

[11] X. Zhou and R. Zafarani, "A survey of fake news," ACM Comput. Surv., vol. 53, no. 5, pp. 1–40, Sep. 2021.

[12] P. Qi et al., "Improving fake news detection by using an entity-enhanced framework to fuse diverse multimodal clues," in Proceedings of the 29th ACM International Conference on Multimedia, New York, NY, USA: ACM, Oct. 2021. doi: 10.1145/3474085.3481548.

[13] B. D. Horne, J. Nørregaard, and S. Adali, "Robust fake news detection over time and attack," ACM Trans. Intell. Syst. Technol., vol. 11, no. 1, pp. 1–23, Feb. 2020.

[14] S. Vosoughi, D. Roy, and S. Aral, "The spread of true and false news online," Science, vol. 359, no. 6380, pp. 1146–1151, Mar. 2018.

[15] M. T. Islam, T. Ahmed, A. B. M. Raihanur Rashid, T. Islam, M. S. Rahman, and M. Tarek Habib, "Convolutional neural network based partial face detection," in 2022 IEEE 7th International conference for Convergence in Technology (I2CT), IEEE, Apr. 2022. doi: 10.1109/i2ct54291.2022.9825259.

[16] T. Islam, A. Kundu, N. Islam Khan, C. Chandra Bonik, F. Akter, and M. Jihadul Islam, "Machine learning approaches to predict breast cancer: Bangladesh perspective," in Smart Innovation, Systems and Technologies, in Smart innovation, systems and technologies. , Singapore: Springer Nature Singapore, 2022, pp. 291–305.

[17] T. Islam et al., "Lexicon and deep learning-based approaches in sentiment analysis on short texts," J. Comput. Commun., vol. 12, no. 01, pp. 11–34, 2024.

[18] U. Roy et al., "Enhancing Bangla fake news detection using bidirectional gated recurrent units and deep learning techniques," in Proceedings of the 7th International Conference on Networking, Intelligent Systems and Security, New York, NY, USA: ACM, Apr. 2024, pp. 1–10.

[19] I. Jahan, M. N. Islam, M. M. Hasan, and M. R. Siddiky, "Comparative analysis of machine learning algorithms for sentiment classification in social media text," World J. Adv. Res. Rev., vol. 23, no. 3, pp. 2842–2852, Sep. 2024.

[20] M. A. Sheakh et al., "Comparative analysis of machine learning algorithms for ECG-based heart attack prediction: A study using Bangladeshi patient data," World J. Adv. Res. Rev., vol. 23, no. 3, pp. 2572–2584, Sep. 2024.

[21] "Enhancing salesforce with machine learning: Predictive analytics for optimized workflow automation," nano-ntp, vol. 20, no. S14, Nov. 2024, doi: 10.62441/nano-ntp.v20is14.91.

[22] T. Islam et al., "Predictive modeling for breast cancer classification in the context of Bangladeshi patients by use of machine learning approach with explainable AI," Sci Rep, vol. 14, no. 1, p. 8487, Apr. 2024.