



Optimizing Misinformation Control: A Cloud-Enhanced Machine Learning Approach



Muhammad Daniyal Baig¹, Waseem Akram², Hafiz Burhan Ul Haq^{3*}, Hassan Zahoor Rajput³,
Muhammad Imran⁴

¹ Department of Computer Science, Lahore Garrison University, 54000 Lahore, Pakistan

² Faculty of Information Technology, University of Central Punjab, 54000 Lahore, Pakistan

³ Department of Electronics and Telecommunication Engineering, Faculty of Engineering, King Mongkut's University of Technology Thonburi (KMUTT), 10140 Bangkok, Thailand

⁴ Department of Economics, Management and Business Law, Università Degli Studi Di Bari Aldo Moro, 70121 Bari, Italy

* Correspondence: Hafiz Burhan Ul Haq (hafiz.burh@kmutt.ac.th)

Received: 12-05-2023

Revised: 01-10-2024

Accepted: 01-20-2024

Citation: M. D. Baig, W. Akram, H. B. U. Haq, H. Z. Rajput, and M. Imran, "Optimizing misinformation control: A cloud-enhanced machine learning approach," *Inf. Dyn. Appl.*, vol. 3, no. 1, pp. 1–11, 2024. <https://doi.org/10.56578/ida030101>.



© 2024 by the authors. Published by Acadlore Publishing Services Limited, Hong Kong. This article is available for free download and can be reused and cited, provided that the original published version is credited, under the CC BY 4.0 license.

Abstract: The digital age has witnessed the rampant spread of misinformation, significantly impacting the medical and financial sectors. This phenomenon, fueled by various sources, contributes to public distress and information warfare, necessitating robust countermeasures. In response, a novel model has been developed, integrating cloud computing with advanced machine learning techniques. This model prioritizes the identification and mitigation of false information through optimized classification strategies. Utilizing diverse datasets for predictive analysis, the model employs state-of-the-art algorithms, including K-Nearest Neighbors (KNN) and Random Forest (RF), to enhance accuracy and efficiency. A distinctive feature of this approach is the implementation of cloud-empowered transfer learning, providing a scalable and optimized solution to address the challenges posed by the vast, yet often unreliable, information available online. By harnessing the potential of cloud computing and machine learning, this model offers a strategic approach to combating the prevalent issue of misinformation in the digital world.

Keywords: Machine learning; Cloud computing; K-Nearest Neighbors (KNN); Random Forest (RF); Data analysis; Misinformation mitigation

1 Introduction

In the contemporary digital landscape, where the internet and social media facilitate rapid information dissemination, the emergence of fake news poses a formidable challenge to credible journalism, public discourse, and the integrity of democratic processes. The widespread dissemination of misinformation, which includes intentionally deceptive content and manipulated narratives, highlights an urgent need for effective mechanisms to distinguish accurate information from falsified stories. With the evolution of traditional news consumption and the rise of online platforms as primary information sources, the development of automated strategies to counteract the proliferation of fake news is increasingly critical. This study explores the application of machine learning techniques in addressing the complex issue of fake news classification. Machine learning, a pivotal subfield of artificial intelligence, demonstrates significant potential in analyzing large data volumes, uncovering hidden patterns, and facilitating informed predictions. The utilization of machine learning in this research represents a key contribution to efforts aimed at curtailing fake news dissemination and enhancing information accuracy.

Fake news encompasses a broad spectrum of deceptive practices, ranging from completely fabricated stories to subtly misleading headlines and selectively edited information. For instance, the spread of false health-related claims, such as unsubstantiated cures for serious ailments, or the circulation of doctored images misrepresenting public figures, exemplifies the multifaceted nature of fake news. As fake news evolves and adapts, an equally dynamic and responsive approach is required to effectively counter it. Linguistic and contextual analysis plays an integral role in fake news identification. Articles characterized by fake news often employ emotionally charged language to

manipulate reader sentiment and may lack direct quotations or references to credible sources, complicating claim verification. Additionally, the temporal and geographical contexts of news stories are critical factors in assessing their authenticity. For example, an emergent news item might initially lack comprehensive corroborative evidence, contrasting with fake news articles presenting grandiose claims devoid of verifiable substantiation.

The complexity of fake news, with its diverse linguistic, contextual, and socio-cultural dimensions, demands a multifaceted analytical approach. This study examines a range of machine learning algorithms and methods capable of discerning between authentic and fabricated news content. Models trained on extensive datasets comprising both genuine and fake news instances are developed, aiming to enhance their capability to generalize from observed patterns and features. These models are designed to identify characteristics commonly associated with unreliable information. Challenges inherent in fake news classification, particularly the evolving strategies used by misinformation disseminators to evade traditional detection, are a focus of discussion in this paper. Furthermore, the potential of various features and data representations in assisting machine learning models to accurately classify content is explored. This includes a critical evaluation of the ethical implications associated with the implementation of automated content filtering systems, emphasizing the need for transparency, bias reduction, and the preservation of user agency.

This research contributes to the burgeoning field focused on mitigating the detrimental impacts of fake news through machine learning applications. By integrating insights from natural language processing, machine learning, and information verification, this work seeks to improve the precision of fake news detection systems, promoting a more informed and discerning digital community. Machine learning techniques offer a viable approach to addressing the complex challenge posed by fake news. Training models on datasets encompassing both authentic and fabricated news articles enables these models to discern patterns, linguistic cues, and contextual features indicative of the reliability of content. For example, a machine learning model might identify that fake news articles frequently utilize sensationalist language, omit credible sources, or present narrative inconsistencies. These identified patterns empower the model to accurately predict the authenticity of new, unexamined articles. The subsequent sections of this article outline the methodology, experimental setup, results, and discussions, culminating in an extensive analysis of the role of machine learning in combating fake news.

2 Related Work

In the realm of fake news detection, a significant expansion of innovative methodologies has been observed. Extensive research has been conducted in the areas of feature extraction, neural network architectures, and sentiment analysis to discern between credible and non-credible information sources. Emphasis has been placed on social network analysis, with numerous studies examining patterns of information dissemination on social platforms to identify potential fake news origins. Additionally, the unique challenges posed by fake news in specific sectors, particularly in healthcare and finance, have been a focus of specialized research. These sectors are notably vulnerable due to the severe implications of misinformation in health and economic decision-making. In response to these challenges, domain-specific datasets have been developed, serving as benchmarks for evaluating the performance of various fake news detection models. Concurrently, the integration of cloud computing with machine learning has attracted considerable attention.

Cloud-based solutions offer both scalability and accessibility, crucial for processing large datasets and facilitating real-time analysis essential in countering the swift spread of fake news. The utilization of cloud infrastructure significantly enhances the computational power necessary for executing intricate machine learning algorithms and managing the vast data prevalent on the internet. Summarily, research efforts in combating fake news have ranged from traditional fact-checking to advanced machine learning innovations. Notably, the application of transfer learning and cloud computing has emerged as influential in augmenting the accuracy and scalability of fake news detection methods. This evolving research landscape highlights the critical need for effective solutions against misinformation in the digital era. The review of literature on text classification for fake news detection reveals a diverse array of machine learning techniques and approaches. This section provides a comprehensive overview of seminal studies in the field, showcasing the evolution and variety of methodologies employed.

In the study conducted by Mujilawati et al. [1], Python-based machine learning was utilized for text classification, underscoring the significance of natural language processing in the realm of fake news detection. Their novel approach, published in the *Journal of Physics: Conference Series*, exemplifies the integration of advanced computing techniques in media analysis. Chang [2] delved into the application of machine learning for text classification to detect "Asian values" in Asian news. This work illustrates the adaptability of text classification techniques to region-specific contexts, highlighting the versatility of machine learning in diverse media landscapes. Zulfarnain and Saqlain [3] presented an extensive survey on text classification algorithms, providing critical insights into state-of-the-art methodologies. Their comprehensive review serves as an invaluable resource for understanding the various approaches and evaluating their effectiveness. Hassan et al. [4] performed analytics on a range of machine learning algorithms for text classification. Their study offers a benchmark for assessing the performance of different methods,

contributing significantly to the field of machine learning in media analysis.

Occhipinti et al. [5] engaged in a comparative analysis of 12 machine learning models for text classification, aiming to identify the most effective techniques for fake news detection. This comparative approach is instrumental in delineating the strengths and weaknesses of various models in the context of media veracity. Ling et al. [6] conducted a case study on the application of interpretable machine learning for text classification in clinical computed tomography reports, demonstrating the potential of these techniques in domain-specific scenarios. Luo [7] explored the efficiency of text classification using selected machine learning techniques. This research contributes to the ongoing dialogue regarding the optimization of classification processes in the context of large data sets. Kadhim [8] executed a survey on supervised machine learning techniques for automatic text classification. This summary provides a comprehensive overview of existing knowledge in the field, highlighting key developments and trends. Blohm et al. [9] evaluated the use of automated machine learning (AutoML) tools for text classification. Their findings underscore the potential synergy between automated processes and human expertise in the domain of text analysis. Janani and Vijayarani [10] applied machine learning and optimization algorithms for automatic text classification, emphasizing the significance of algorithm selection in achieving precise and accurate results.

Tong and Koller [11] investigated the application of support vector machine (SVM) active learning to text classification, contributing significantly to the field of active learning methodologies. In the specific arena of fake news detection, several researchers have made notable contributions. Surekha et al. [12] combined the Web of Trust (WoT) with Asian social networks to enhance feature extraction and the classification of text and images, thereby improving the detection of digital misinformation and fake news. Kurasinski and Mihailescu [13] underscored the crucial role of explainability in text classification for fake news detection, providing valuable insights into making the decision-making process more transparent. Bangyal et al. [14] employed deep learning techniques to detect fake news related to COVID-19, underlining the significance of this issue in the context of public health crises. Madhuravani et al. [15] proposed a diverse approach to fake news classification utilizing machine learning, reflecting the wide range of methodologies addressing this challenge. Dubey et al. [16] developed a framework for fake news classification that integrates vectorization with machine learning techniques, offering a novel approach that marries text analysis with advanced computational methods. Ahmed et al. [17] focused on an explainable text classification model tailored for COVID-19 fake news detection, addressing the increasing demand for interpretability in machine learning applications.

Roshinta et al. [18] compared various text classification methods for detecting fake news on Indonesian websites, emphasizing the necessity of accommodating regional and linguistic variations in fake news detection models. Aljabri et al. [19] contributed to the field by utilizing machine learning models for fake news detection, highlighting the pressing nature of this issue. Gururaj et al. [20] proposed a machine learning-based approach for fake news detection, underscoring the multidisciplinary nature of this research area. Hasan and Itu [21] explored innovative strategies for detecting fake news using machine learning, illustrating the dynamic and ever-evolving landscape of this field. This literature review collectively demonstrates the burgeoning interest and diverse methodologies in the domain of text classification for fake news detection. It highlights the interdisciplinary nature of this research, emphasizing its critical role in tackling misinformation and disinformation within society. Researchers continue to refine and adapt their methodologies, confronting the challenges posed by the evolving landscape of fake news.

Riaz et al. [22] have innovatively employed machine learning for the detection of fake news, underscoring the critical role of these techniques in distinguishing between reliable and unreliable information sources. Their research, published in the *International Journal of Innovative Technology and Exploring Engineering*, elucidates the significance of natural language processing in the realm of fake news detection. This work exemplifies the application of machine learning in media analysis, offering a novel perspective in the field. In a distinct context, Haq and Saqlain [23] responded to the challenges posed by the COVID-19 pandemic by developing an iris detection system for monitoring attendance in educational institutes. Their application of machine learning presents a pragmatic approach to verifying student presence during the pandemic. This study, featured in the *Journal of Industrial Intelligence*, not only demonstrates the versatility of machine learning but also its potential to address pressing real-world issues. Furthermore, Baig et al. [24] investigated the use of cloud gaming in the learning of programming concepts. While their research, published in *Artificial Intelligence and Applications*, does not directly pertain to fake news detection, it highlights the importance of cloud computing and machine learning in enhancing educational methodologies. The integration of cloud computing, offering scalability and accessibility, is essential for processing extensive datasets and conducting real-time analysis. These attributes are also beneficial in the context of fake news detection, indicating the wide applicability of cloud computing and machine learning across various domains.

Nawaz et al. [25] conducted an in-depth review of cloud computing services, emphasizing the security challenges inherent in such systems. Their work, published in the *Lahore Garrison University Research Journal of Computer Science and Information Technology*, underscores the criticality of addressing security concerns. This aspect is particularly pertinent in the realm of information authenticity, a key concern in the age of fake news. Haq and Saqlain [26] focused their research on deploying effective machine learning strategies to identify Sybil attacks within

IoT networks. Although the context differs, the principles of anomaly detection and network security, as applied in their work, intersect with the domain of fake news detection. Machine learning techniques, in this regard, can be adapted for identifying misinformation propagation patterns in digital platforms. Saqlain [27] made a noteworthy contribution to sustainable hydrogen production using decision-making approaches. While not directly related to fake news, this study exemplifies the power of data-driven decision-making, a core component of machine learning-based fake news detection strategies. In a distinct application, Abid and Saqlain [28] utilized linear programming for optimizing the transportation of bakery products. Although specific to a different field, the principles of linear programming and optimization techniques are crucial in enhancing information dissemination and network behavior analysis. These techniques bear relevance to the identification of patterns and sources in fake news dissemination. Lastly, Jafar et al. [29] leveraged intuitionistic fuzzy soft matrices-based algorithms to improve diabetes diagnosis. Though unrelated to fake news, their research demonstrates the application of advanced computational methods in enhancing diagnostic precision. Similarly, machine learning methods can be employed to augment the accuracy in identifying and verifying fake news sources.

3 Experimental Setup

The experimental setup utilized a computing system equipped with 12 GB RAM, a Core i5, 7th Generation processor, and an RX 580 8GB graphical processing unit. For data gathering and downloading of the dataset, an internet connection with a speed of 10 MBPS was employed. The Anaconda Jupyter Notebook platform was used for conducting the experiments.

3.1 Datasets

The datasets used in this study are open-source and publicly available, encompassing major classes of "fake" and "truth" news, derived from multiple fields to introduce diversity. The authenticity of the datasets was verified using established fake news authenticity websites and fact-checking platforms like snopes.com. The ISOT Fake News dataset is an open-source collection crafted to address the issue of fake news and misinformation in the media landscape. This dataset, primarily consisting of textual content such as news articles and headlines, serves as a crucial resource for developing algorithms capable of detecting misleading or deceptive information. Each entry in the dataset is labeled as either "fake" or "real," facilitating the training and evaluation of machine learning models. With its broad spectrum of sources covering various domains and subjects, the dataset captures the complexity of online content. The ISOT Fake News dataset serves as a pivotal tool for researchers engaged in the exploration of innovative methodologies in the domain of fake news detection. This dataset facilitates the analysis of misinformation dissemination patterns and plays a significant role in the global endeavor to enhance media literacy and ensure the sharing of accurate information.

Similarly, the DS2 Fake News dataset is a meticulously curated collection, specifically designed to address the widespread issue of fake news and misinformation. Encompassing a wide range of textual content, including news articles and headlines, this dataset stands as an invaluable resource for researchers and practitioners focused on developing and refining algorithms to identify deceptive or false information. Each content piece within this dataset is distinctly categorized as "fake" or "real", offering a well-labeled basis for the application and assessment of machine learning models. The variety in sources and types of content in the DS2 Fake News dataset is intended to encapsulate the complexity and multidimensionality of misleading information encountered in digital media. This dataset is instrumental in empowering researchers to investigate new methods, techniques, and approaches in the field of fake news detection, thereby contributing significantly to the overarching goal of promoting information accuracy and media literacy.

4 Discussion

This section delineates the methodology employed in the proposed model, which integrates cloud computing with the amalgamation of datasets.

4.1 Proposed Model

The framework described herein comprises a fake news detection system, enhanced through the integration of cloud computing. This system is an assemblage of existing state-of-the-art methodologies. The model uniquely combines both supervised and unsupervised learning algorithms to forge an efficient system capable of achieving high accuracy and real-time results. The novelty of this framework lies in the fusion of cloud computing with these learning algorithms. The input data for this model consists of both numerical and qualitative elements. Variables such as the writer's name, publication date, and content (represented as a collection of words) form the basis of the input stream. These variables are utilized in transfer learning and cloud computing integration. Preprocessing of the dataset is conducted to filter, clean, and remove outliers, as machine learning models are particularly sensitive to

such anomalies. The data is then partitioned into three distinct sets: training, testing, and validation, to facilitate the experimentation process. Figure 1 illustrates the proposed model.

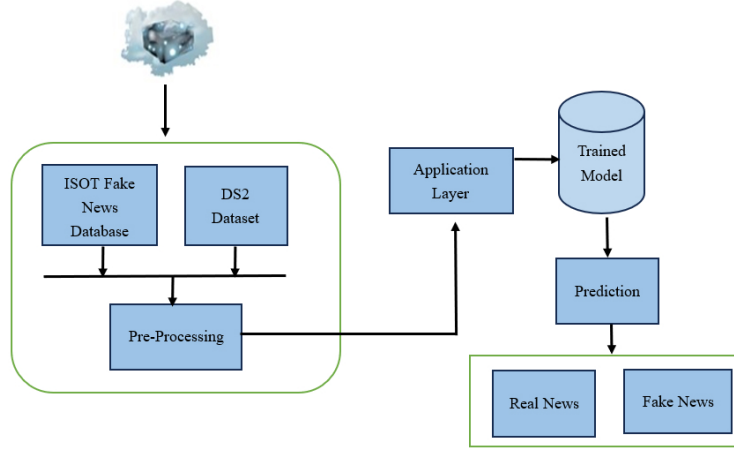


Figure 1. Proposed model

In the context of binary classification using machine learning algorithms, a general mathematical model is outlined. This model encapsulates the fundamental components and equations necessary for constructing and assessing a binary classifier. The dataset, denoted as $\{(x_{-1}, y_{-1}), (x_{-2}, y_{-2}), \dots, (x_{-n_0}, y_{-n})\}$, comprises feature vectors x_i for each instance i , with corresponding binary labels y_i (either 0 or 1) indicating the class.

The objective of a binary classifier is to ascertain a function $h(x)$, mapping the feature vector x to a predicted label y . The decision boundary of the classifier is defined by a set of parameters θ . The model can be represented as:

$$[[y' = Iy\{\text{Classifier}\}(x; \theta)] \quad (1)$$

4.1.1 Loss Function

In the employed model, the loss function serves to quantify the error between the predicted label (x) and the actual label (y). For the purpose of binary classification, the model incorporates standard loss functions such as the logistic loss, also known as cross-entropy, suitable for probabilistic classifiers like logistic regression. Additionally, the hinge loss function is utilized for SVM.

$$L(y', y) = -[y \cdot \log(y') + (1 - y) \cdot \log(1 - y')] \quad (2)$$

The objective function to be minimized during the training process is the aggregate loss across all instances in the hinge loss.

$$L(y', y) = \max(0, 1 - y' \cdot y) \quad (3)$$

$$J(\theta) = \frac{1}{n} \sum_{i=1}^n L(\text{Classifier}(x_i; \theta), y_i) \quad (4)$$

4.1.2 Training

During the training phase, the focus is on determining the optimal parameter values, denoted as θ , that minimize the objective function. This optimization is realized through various techniques, including gradient descent and other algorithms specifically designed for certain classifiers.

4.1.3 Prediction

For the prediction of the class of a new instance, referred to as x_{news} , the trained model is applied to calculate the predicted label, as depicted in Eq. (5).

$$y'_{\text{new}} = \text{Classifier}(x_{\text{new}}; \theta) \quad (5)$$

In this binary classification framework, the final prediction is based on a threshold, set at a decision boundary, which, for the purposes of this study, is established at 0.7.

4.2 Preprocessing of Dataset

The dataset comprises three types of data: textual, numerical, and categorical. Textual data encompasses the content of the news, whereas categorical data includes variables such as the author’s name. Numerical data refers to details like the date of publication.

4.3 Machine Learning Models

Six algorithms have been utilized in the proposed model: RF, KNN, and Decision Tree (DT) for the classification of news as either fake or real. Furthermore, Gradient Boosting (GBoost) and Extreme GBoost (XGBoost) are implemented to enhance the accuracy and reduce the misclassification rate of the model.

4.4 Testing, Training & Validation

The datasets, sourced from repositories such as UCI and Kaggle, consist of a blend of authentic and fabricated news items, facilitating a qualitative analysis. Two distinct datasets are employed to evaluate the performance of the proposed model comprehensively. The ISOT Fake News dataset as shown in Table 1, primarily used for binary classification, includes 21,415 instances of genuine news and 23,481 instances of fake news. This dataset spans various categories, including world news, political news, government news, and Middle East news, among others. The DS2 Fake News dataset as illustrated in Table 2 is segmented into two primary classes: fake news and real news, comprising 37,115 instances of fake news and 66,067 instances of genuine news. These datasets serve to strengthen the performance assessment of the model, offering an extensive view of the detection and categorization of fake news.

Table 1. ISOT fake news dataset

Label Name	Data Points
True	21,415
False	23,481

Table 2. DS2 dataset

Label Name	Data Points
True	66,067
False	37115

4.5 Machine Learning Algorithms

RF, an ensemble learning technique, is frequently employed in text classification. This method aggregates predictions from numerous DTs, categorizing documents by their content. RF selects subsets of data and features randomly for each tree to mitigate overfitting. The final classification decision is derived from either majority voting or averaging the predictions of these trees, thereby enhancing performance, especially for high-dimensional text data. KNN is utilized for its proximity-based class assignment in text classification. It represents documents as feature vectors and determines the class by identifying the most common label among the K nearest neighbors in the training dataset. KNN is particularly effective for complex boundaries in text data; however, its performance is contingent upon the chosen distance metrics, the value of K, and the representation of the data. DTs play a pivotal role in text classification, making binary decisions based on textual features and constructing a tree-like structure. While they effectively capture patterns and handle various feature types, DTs are susceptible to overfitting. Techniques such as pruning and ensembling are often implemented to enhance their utility in text classification. GBoost, another ensemble learning method, combines predictions from multiple weak models. This technique iteratively refines its predictions, focusing on correcting the errors of preceding models, and is widely recognized for its efficacy in text classification tasks. Naive Bayes, a probabilistic classifier grounded in Bayes’ theorem, operates under the assumption of feature independence. Despite its simplicity, Naive Bayes is remarkably effective in text classification and spam filtering, owing to its capability to handle large feature spaces efficiently. SVMs are robust classifiers that

delineate the optimal hyperplane to separate different classes. SVMs excel in maximizing the margin between data points and can accommodate both linear and nonlinear boundaries through the application of kernel functions.

5 Experimental Analysis

This section delineates the experimental phase of the study.

5.1 Evaluation Metrics

The outcomes of the training phase and the results yielded by the proposed model, employing transfer learning, are discussed in this segment. The primary objective of this research was to devise a swift and efficient mechanism for differentiating between authentic and fabricated news. The experiments were conducted using the Anaconda Jupyter Notebook environment, focusing on binary classification. The computational process was executed on a single RX 580 with 8 GB of available RAM. The dataset used in this study was apportioned into 80% for training and 20% for validation. Various performance parameters such as sensitivity, specificity, precision, and accuracy were employed to assess the model's efficacy.

$$sensitivity = \frac{\left(\frac{T_p}{E_p}\right)}{\left(\frac{T_p}{t_p}\right) + \left(\frac{T_m}{t_m}\right)} * 100 \quad (6)$$

$$specificity = \frac{\left(\frac{T_m}{E_m}\right)}{\left(\frac{T_m}{t_m}\right) + \left(\frac{T_e}{t_e}\right)} * 100 \quad (7)$$

$$precision = \frac{\left(\frac{T_p}{E_p}\right)}{\left(\frac{T_p}{t_p}\right) + \left(\frac{T_e}{t_e}\right)} * 100 \quad (8)$$

$$accuracy = \frac{\left(\frac{T_p}{t_p}\right) + \left(\frac{T_m}{t_m}\right)}{p + m} * 100 \quad (9)$$

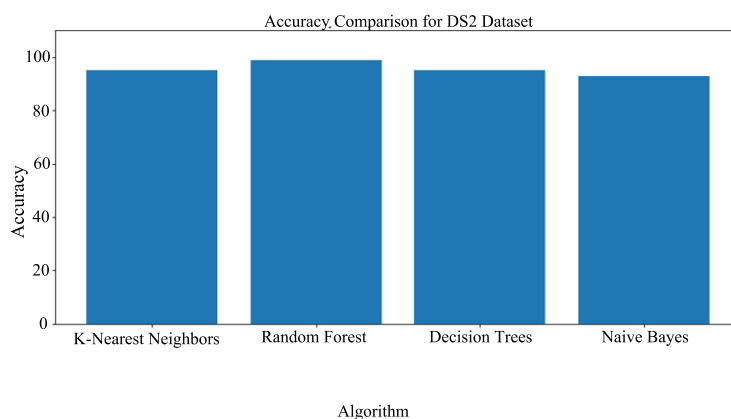
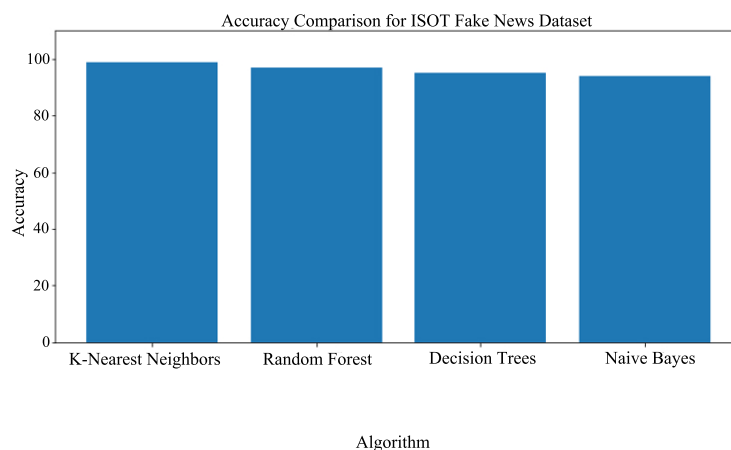
Table 3 encapsulates the overall results of the experimental phase. The proposed model was applied to two distinct datasets: the ISOT Fake News dataset and the DS2 dataset. The results underscore the effectiveness of a range of machine learning algorithms across diverse datasets. Notable findings include the KNN algorithm achieving an exceptional accuracy of 99% on the DS1 dataset, indicating its proficiency in that context. The RF algorithm exhibited robust performance with an accuracy of 97%, affirming its reliability. DTs achieved an accuracy of 95%, demonstrating their capability, while Naive Bayes also showed substantial potential, achieving an accuracy of 94%. Collectively, these results highlight the adaptability of the algorithms to various datasets and underline their individual strengths. The analysis of the DS2 dataset reveals interesting patterns in the performance of different algorithms.

The KNN algorithm exhibited versatility, achieving an accuracy of 95%, indicative of its capacity for precise predictions. The RF algorithm emerged as a standout, attaining an impressive accuracy of 99%, thereby highlighting its ability to navigate complex relationships within the data. Notably, the Naive Bayes algorithm achieved an accuracy of 93%, affirming its efficacy even with its relatively straightforward assumptions. These results collectively underscore the diversity and strengths of these algorithms, each excelling in different aspects and contributing to a comprehensive understanding of the datasets' intricacies.

The results from the overall experimentation phase highlight the effectiveness of the proposed model across two distinct datasets: the ISOT Fake News dataset and the DS2 dataset. In the ISOT Fake News dataset, KNN demonstrated exceptional accuracy, achieving 99%, showcasing its proficiency in DS1. Concurrently, RF displayed strong performance with an accuracy of 97%, positioning itself as a robust contender in the DS1 context. DTs exhibited notable competence with an accuracy of 95%, while Naive Bayes showed potential with an accuracy of 94%. Transitioning to the DS2 dataset, KNN maintained its prowess with an accuracy of 95%, reinforcing its capabilities. RF sustained its high-performance trajectory with an accuracy of 99%, illustrating its robustness in DS2. DTs also maintained their competency with an accuracy of 95% in DS2. These findings illustrate the diversity of algorithmic strengths across different datasets, highlighting both proficiency and robustness in various contexts. Figures 2 and 3 provide visual representations of the DS2 Dataset and ISOT Dataset, respectively.

Table 3. Dataset proficiency

Dataset	Algorithm	Accuracy	Description
ISOT Fake News dataset	KNN	99%	Proficiency in DS1
ISOT Fake News dataset	RF	97%	Strong performance in DS1
ISOT Fake News dataset	DT	95%	Notable competence in DS1
ISOT Fake News dataset	Naive Bayes	94%	Potential shown in DS1
DS2 dataset	KNN	95%	Notable competence in DS2
DS2 dataset	RF	99%	Robust performance in DS2
DS2 dataset	DT	95%	Notable competence in DS2

**Figure 2.** DS2 dataset**Figure 3.** ISOT dataset**Table 4.** Experimental analysis

Dataset	Algorithm	Precision Estimate	Sensitivity Estimate	Specificity Estimate
ISOT Fake News dataset	KNN	99%	94%	96%
ISOT Fake News dataset	RF	97%	89%	85%
ISOT Fake News dataset	DT	95%	91%	92%
ISOT Fake News dataset	Naive Bayes	94%	92%	91%
DS2 dataset	KNN	95%	93%	99%
DS2 dataset	RF	99%	97%	92%
DS2 dataset	DT	95%	96%	94%
DS2 dataset	Naive Bayes	91%	92%	90%

Table 4 provides a comprehensive evaluation of various algorithms' performance across two distinct datasets. For the ISOT Fake News Dataset, it was observed that KNN achieved a high precision of 99%, complemented by

a sensitivity of 94% and a specificity of 96%. RF demonstrated a strong precision of 97%, alongside a sensitivity of 89% and a specificity of 85%. DTs exhibited a balanced performance with a precision of 95%, a sensitivity of 91%, and a specificity of 92%. Naive Bayes, meanwhile, showed competitive precision at 94%, with a sensitivity of 92% and specificity of 91%. Transitioning to the DS2 Dataset, KNN maintained notable precision at 95%, coupled with a sensitivity of 93% and a high specificity of 99%. RF, in this dataset, showcased robust precision at 99%, with sensitivity and specificity estimates of 97% and 92%, respectively. DTs sustained their precision at 95%, achieving a sensitivity of 96% and a specificity of 94%. Naive Bayes in the DS2 Dataset demonstrated a precision of 91%, accompanied by sensitivity and specificity estimates of 92% and 90%, respectively. These findings collectively highlight the algorithms' efficacy, underscoring their strengths and limitations in terms of precision, sensitivity, and specificity across different datasets. Figure 4 illustrates a line graph representation of the performance of these machine learning models.

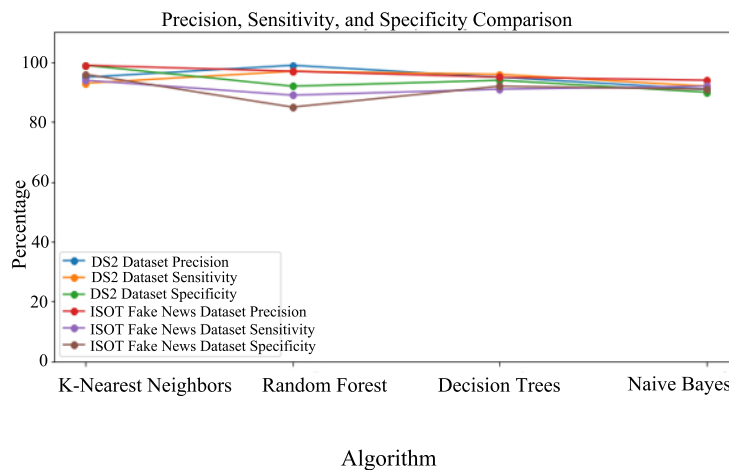


Figure 4. Line graph of machine learning models

6 Conclusion

The extensive analysis conducted in this study provides insightful evaluations of various algorithms' performance across two distinct datasets. In the ISOT Fake News dataset, the proficiency of the algorithms is manifested in their precision, sensitivity, and specificity. KNN achieved a notable precision of 99%, supplemented by a sensitivity of 94% and specificity of 96%. RF demonstrated substantial precision at 97%, albeit with marginally lower sensitivity (89%) and specificity (85%). DT presented a well-balanced performance, evidenced by their 95% precision, 91% sensitivity, and 92% specificity. Naive Bayes, meanwhile, exhibited commendable precision at 94%, along with sensitivity and specificity of 92% and 91%, respectively. Transitioning to the DS2 dataset, KNN maintained its noteworthy precision at 95%, accompanied by a sensitivity of 93% and an impressive specificity of 99%. RF continued its strong performance trajectory, achieving a robust precision of 99%, paired with sensitivity and specificity estimates of 97% and 92%, respectively. DT in the DS2 dataset sustained their precision at 95%, while recording a sensitivity of 96% and specificity of 94%. Naive Bayes, within this dataset, demonstrated a precision of 91%, along with sensitivity and specificity estimates of 92% and 90%, respectively. These results collectively underscore the adaptability and unique strengths of the diverse algorithms in handling different datasets. They offer a comprehensive understanding of the algorithms' performance nuances in terms of precision, sensitivity, and specificity. This research thereby contributes significantly to the field of fake news detection, providing valuable insights into the efficacy of various machine learning algorithms in different contexts.

Data Availability

The data used to support the research findings are available from the corresponding author upon request.

Conflicts of Interest

The authors declare no conflict of interest.

References

- [1] S. Mujilawati, M. Sholihin, R. Wardhani, and M. R. Zamroni, "Python based machine learning text classification," *J. Phys.: Conf. Ser.*, vol. 2394, no. 1, p. 012015, 2022. <https://doi.org/10.1088/1742-6596/2394/1/012015>

- [2] L. Chang, "Detecting Asian values in Asian news via machine learning text classification," *Adv. Data Sci. Inform. Eng.*, pp. 123–128, 2021.
- [3] M. Zulqarnain and M. Saqlain, "Text readability evaluation in higher education using CNNs," *J. Ind. Intell.*, vol. 1, no. 3, pp. 184–193, 2023. <https://doi.org/10.56578/jii010305>
- [4] S. U. Hassan, J. Ahamed, and K. Ahmad, "Analytics of machine learning-based algorithms for text classification," *Sustain. Oper. Comput.*, vol. 3, pp. 238–248, 2022.
- [5] A. Occhipinti, L. Rogers, and C. Angione, "A pipeline and comparative study of 12 machine learning models for text classification," *Expert Syst. Appl.*, vol. 201, p. 117193, 2022. <https://doi.org/10.48550/arXiv.2204.06518>
- [6] T. Ling, L. Jake, J. Adams, K. Osinski, X. Liu, and D. Friedland, "Interpretable machine learning text classification for clinical computed tomography reports – A case study of temporal bone fracture," *Comput. Methods Programs Biomed. Update*, 2023. <https://doi.org/10.1016/j.cmpbup.2023.100104>
- [7] X. Luo, "Efficient english text classification using selected machine learning techniques," *Alex. Eng. J.*, vol. 60, pp. 3401–3409, 2021. <https://doi.org/10.1016/j.aej.2021.02.009>
- [8] A. I. Kadhim, "Survey on supervised machine learning techniques for automatic text classification," *Artif. Intell. Rev.*, vol. 52, pp. 273–292, 2019. <https://doi.org/10.1007/s10462-018-09677-1>
- [9] M. Blohm, M. Hanussek, and M. Kintz, "Leveraging automated machine learning for text classification: Evaluation of AutoML tools and comparison with human performance," in *International Conference on Agents and Artificial Intelligence*, 2020, pp. 1131–1136. <https://doi.org/10.5220/0010331411311136>
- [10] R. Janani and S. Vijayarani, "Automatic text classification using machine learning and optimization algorithms," *Soft Comput.*, vol. 25, pp. 1129–1145, 2020. <https://doi.org/10.1007/s00500-020-05209-8>
- [11] S. Tong and D. Koller, "SVM active learning with applications to text classification," *J. Mach. Learn. Res.*, vol. 2, pp. 45–66, 2022.
- [12] T. L. Surekha, N. C. S. Rao, C. K. Shahnazeer, S. M. Yaseen, S. K. Shukla, S. Bharat, and M. Arumugam, "Digital misinformation and fake news detection using WoT integration with asian social networks fusion based feature extraction with text and image classification by machine learning architectures," *Theor. Comput. Sci.*, vol. 927, pp. 1–14, 2022. <https://doi.org/10.1016/j.tcs.2022.05.017>
- [13] L. Kurasinski and R. Mihailescu, "Towards machine learning explainability in text classification for fake news detection," in *19th IEEE International Conference on Machine Learning and Applications (ICMLA)*. Miami, FL, USA, 2020, pp. 775–781. <https://doi.org/10.1109/ICMLA51294.2020.00127>
- [14] W. H. Bangyal, R. Qasim, N. U. Rehman, Z. Ahmad, H. S. Dar, L. Rukhsar, Z. Aman, and J. Ahmad, "Detection of fake news text classification on COVID-19 using deep learning approaches," *Comput. Math. Methods Med.*, pp. 1–14, 2021. <https://doi.org/10.1155/2021/5514220>
- [15] K. Madhuravani, N. Vamshika, B. Akhila, P. V. Kumar, and V. S. Reddy, "Fake news classification model using machine learning," *YMER Digital*, 2022.
- [16] Y. Dubey, P. Wankhede, A. Borkar, T. Borkar, and P. Palsodkar, "Framework for fake news classification using vectorization and machine learning," *Stud. Comput. Intell.*, pp. 327–343, 2021.
- [17] M. Ahmed, M. Hossain, R. Islam, and K. Andersson, "Explainable text classification model for COVID-19 fake news detection," *J. Internet Serv. Inf. Secur.*, vol. 12, pp. 51–69, 2022. <https://doi.org/10.1022667/JISIS.2022.05.31.051>
- [18] T. Roshinta, Hartatik, E. Fauziyah, I. Dinata, N. Firdaus, and F. A'la, "A comparison of text classification methods: Towards fake news detection for Indonesian websites," in *2022 1st International Conference on Smart Technology, Applied Informatics, and Engineering (APICS)*. Surakarta, Indonesia, 2022, pp. 59–64. <https://doi.org/10.1109/APICS56469.2022.9918702>
- [19] M. Aljabri, D. Alomari, and M. Aboulmour, "Fake news detection using machine learning models," in *14th International Conference on Computational Intelligence and Communication Networks (CICN)*. Al-Khobar, Saudi Arabia, 2022, pp. 473–477. <https://doi.org/10.1109/CICN56167.2022.10008340>
- [20] H. Gururaj, H. Lakshmi, B. C. Soundarya, F. Flammuni, and V. Janhavi, "Machine learning-based approach for fake news detection," *J. ICT Stand.*, vol. 10, pp. 509–530, 2022. <https://doi.org/10.13052/jicts2245-800X.1042>
- [21] M. Hasan and I. Itu, "A distinctive approach for detecting fake news using machine learning," *Int. J. Innov. Technol. Explor. Eng.*, 2022. <https://doi.org/10.35940/ijitee.C9640.0111322>
- [22] M. Riaz, A. Habib, M. Saqlain, and M. S. Yang, "Cubic bipolar fuzzy-VIKOR method using new distance and entropy measures and Einstein averaging aggregation operators with application to renewable energy," *Int. J. Fuzzy Syst.*, vol. 25, no. 2, pp. 510–543, 2023. <https://doi.org/10.1007/s40815-022-01383-z>
- [23] H. B. U. Haq and M. Saqlain, "Iris detection for attendance monitoring in educational institutes amidst a pandemic: A machine learning approach," *J. Ind. Intell.*, vol. 1, no. 3, pp. 136–147, 2023. <https://doi.org/10.56578/jii010301>

- [24] D. Baig, W. Akram, H. B. U. Haq, and M. Asif, "Cloud gaming approach to learn programming concepts," *Artif. Intell. Appl.*, 2022. <https://doi.org/10.47852/bonviewAIA32021378>
- [25] S. Nawaz, A. Akhtar, and H. B. U. Haq, "Cloud computing services and security challenges: A review," *Lahore Garr. Univ. Res. J. Comput. Sci. Inf. Technol.*, vol. 7, no. 2, pp. 17–28, 2023. <https://doi.org/10.54692/lgurjcsit.2023.0702459>
- [26] H. B. U. Haq and M. Saqlain, "An implementation of effective machine learning approaches to perform Sybil Attack Detection (SAD) in IoT network," *Theor. Appl. Comput. Intell.*, vol. 1, no. 1, pp. 1–14, 2023.
- [27] M. Saqlain, "Sustainable hydrogen production: A decision-making approach using VIKOR and intuitionistic hypersoft sets," *J. Intell. Manag. Decis.*, vol. 2, no. 3, pp. 130–138, 2023. <https://doi.org/10.56578/jimd020303>
- [28] M. Abid and M. Saqlain, "Decision-making for the bakery product transportation using linear programming," *Spectr. Eng. Manag. Sci.*, vol. 1, no. 1, pp. 1–12, 2023.
- [29] M. N. Jafar, K. Muniba, and M. Saqlain, "Enhancing diabetes diagnosis through an intuitionistic fuzzy soft matrices-based algorithm," *Spectr. Eng. Manag. Sci.*, vol. 1, no. 1, pp. 73–82, 2023.