

Fake News Detection Using Machine Learning Models: A Comparative Study of ISOT and LIAR Datasets

Rudra Mogre

Mountain House High School, 1090 S. Central Parkway, Mountain House, CA 95391, United States

ABSTRACT

The proliferation of fake news across online platforms poses a growing challenge to information integrity and public trust. Traditional fact-checking mechanisms are often too slow to counteract the viral spread of misinformation. As a result, automated approaches using machine learning have emerged as effective tools for distinguishing between legitimate and fabricated news articles. This study applies supervised machine learning techniques to classify news content using the ISOT (~44,898 articles) and LIAR (12,836 statements) benchmark datasets. Feature extraction was performed using Term Frequency-Inverse Document Frequency (TF-IDF) and n-gram analysis (unigrams and bigrams). Five classical algorithms were evaluated: Logistic Regression, Naïve Bayes, Support Vector Machines (SVM), Decision Trees, and Random Forests. Model performance was assessed using accuracy, precision, recall, F1-score, and AUC. On the ISOT dataset, Random Forest achieved the highest accuracy of 0.998, followed by SVM and Decision Tree at 0.996, Logistic Regression at 0.993, and Naïve Bayes at 0.949. On the LIAR dataset, accuracies ranged from 0.571 (Decision Tree) to 0.628 (Logistic Regression). The results demonstrate that classical machine learning models, coupled with robust text representation techniques, can effectively detect misinformation on structured full-article datasets like ISOT while maintaining transparency and scalability, though performance drops significantly on short-claim datasets like LIAR due to structural differences.

Keywords: Fake News Detection; Machine Learning; Text Classification; TF-IDF; Misinformation; Data Mining

INTRODUCTION

The rapid expansion of social media and digital journalism has made information more accessible than ever (1). Still, it has also facilitated the spread of fake news, fabricated content designed to mislead readers or manipulate public perception (2, 3). Fake news has been

linked to political polarization, erosion of trust in media, and the amplification of social divisions (4). Unlike genuine news reporting, fake articles often rely on sensationalist language, emotional framing, and unverified claims (5). Detecting such misinformation manually is infeasible at scale, motivating the use of automated detection systems based on machine learning (6).

Machine learning methods excel at identifying subtle statistical and linguistic cues that differentiate real from fake content (7–10). Prior studies have achieved remarkable accuracy, often above 95% using algorithms such as Logistic Regression, Support Vector Machines, and Naïve Bayes when trained on curated datasets (11). By transforming textual content into numerical features,

Corresponding author: Rudra Mogre, E-mail: rudramogre2009@gmail.com.

Copyright: © 2026 Rudra Mogre. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Accepted January 29, 2026

<https://doi.org/10.70251/HYJR2348.41519524>

models can learn patterns indicative of deception, biased phrasing, or factual inconsistency (12–15). Unlike deep learning methods, which require large datasets and substantial computational power, traditional ML approaches are interpretable, lightweight, and easily adaptable to new domains (16).

This research, therefore, focuses on developing and evaluating classical supervised machine learning models for fake news detection (17). Through rigorous experimentation using TF-IDF features and various classifiers, the study aims to identify the optimal balance between performance and explainability.

METHODS AND MATERIALS

Dataset and Preprocessing

The research utilizes the ISOT Fake News Dataset, which comprises 44,898 labeled news articles (~21,417 true articles from Reuters and ~23,481 fake articles from various flagged sources) categorized as real or fake. Articles were primarily collected from 2016 to 2017, with a focus on political and world news topics. The LIAR dataset consists of 12,836 short, manually labeled factual claims (statements) from PolitiFact.com, rather than full news articles, representing a different data structure (micro-statements vs. complete articles). Before modeling, the data undergoes a series of preprocessing steps to remove noise and ensure consistent representation (Figure 1).

Text preprocessing involved converting all text to lowercase, removing punctuation, stopwords (using scikit-learn's English stopword list), and special

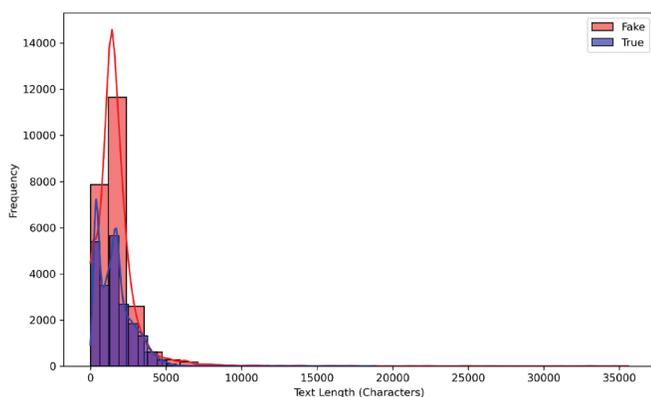


Figure 1. Distribution of text lengths for fake and true news articles in the ISOT dataset, illustrating differences in average length and frequency distributions (sample size: 44,898 articles).

characters, and applying lemmatization (using NLTK's WordNetLemmatizer). Tokenization was performed using the default tokenizer in scikit-learn's CountVectorizer and TfidfVectorizer. No semantic embeddings or deep NLP methods were employed; the approach focused on frequency-based representations (18). The dataset was split into training and testing subsets using an 80:20 ratio.

Feature extraction is conducted using TF-IDF vectorization and CountVectorizer with n-grams (unigrams and bigrams) to capture local linguistic patterns (19). These methods represent each document as a weighted vector of word frequencies, emphasizing distinctive terms across the corpus (20). The resulting feature matrix provides a robust numerical foundation for machine learning algorithms.

Machine Learning Models

Five classical supervised learning algorithms were implemented and compared using the scikit-learn library (version 1.3 or the latest stable release available at the time of experimentation). Logistic Regression was used as a baseline linear model well suited for binary text classification, modeling the probability of class membership through the logistic function and optimized with L2 regularization, where the regularization parameter (C) was evaluated in the range ([0.1, 1, 10]) and the optimal value was (C = 1.0). Naïve Bayes (MultinomialNB) was employed as a probabilistic classifier based on Bayes' theorem, particularly effective for high-dimensional sparse text representations derived from word frequencies. A Support Vector Machine (SVM) with a linear kernel was used to maximize the margin between classes in high-dimensional feature space; the regularization parameter (C) was similarly tested in ([0.1, 1, 10]), with an optimal value of (C = 1.0), providing strong generalization performance. A Decision Tree classifier was included as a non-parametric model that constructs hierarchical decision rules based on feature thresholds, offering interpretability at the cost of increased susceptibility to overfitting. Finally, a Random Forest classifier was implemented as an ensemble method combining multiple decision trees to reduce variance and improve predictive accuracy, with the number of estimators evaluated in ([50, 100, 200]) and an optimal configuration of 100 trees with no maximum depth constraint.

Each model is trained using the preprocessed TF-IDF features. Hyperparameter tuning is performed via grid search and k-fold cross-validation to identify optimal

configurations for parameters such as regularization strength (C), kernel type, and tree depth. The models are implemented using the scikit-learn library in Python, ensuring reproducibility and transparency.

Evaluation Metrics

Performance is assessed using accuracy, precision, recall, and F1-score, with a focus on minimizing false-positive cases where legitimate news is misclassified as fake. A confusion matrix is constructed for each model to visualize classification outcomes, and Receiver Operating Characteristic (ROC) curves with Area Under the Curve (AUC) values are used to evaluate discriminative ability. Cross-validation ensures that results generalize beyond a single train-test split.

RESULTS

The results indicate that Logistic Regression and Linear SVM achieved high accuracy and stability on the ISOT dataset when trained on TF-IDF features. Random Forest attained the highest accuracy of 0.998 on ISOT (Figure 2 and Table 1). On the LIAR dataset, performance was considerably lower across all models (Figure 8), likely due to the short length and different structure of the statements (6). The Naïve Bayes classifier typically performs slightly lower (93–95%) but provides rapid training and strong interpretability. Random Forests may perform competitively but with a higher computational cost, while Decision Trees are shown to overfit due to data sparsity.

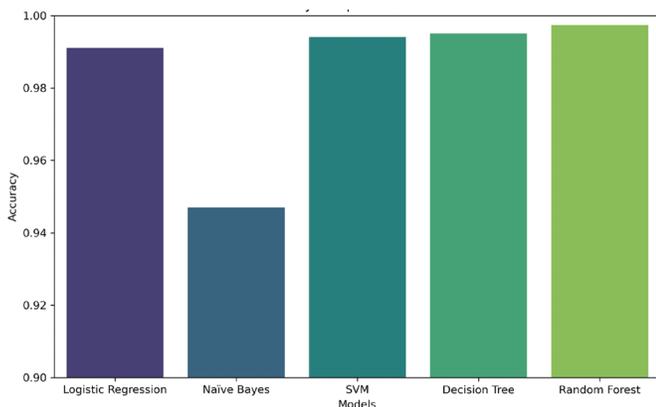


Figure 2. Different machine learning models and their respective accuracies, with Random Forest having the highest accuracy (99.8%) for the ISOT Dataset (44,898 articles).

Table 1. Accuracy of the different models on the Information Security Office of Technology, ISOT, (~44,898 articles) and LIAR (12,836 statements) datasets.

Model	Accuracy ISOT	Accuracy LIAR
Logistic Regression	0.993	0.628
Naïve Bayes	0.949	0.623
Support Vector Machine (SVM)	0.996	0.626
Decision Tree	0.996	0.571
Random Forest	0.998	0.625

Model comparisons will be summarized through accuracy tables and ROC curves. It is shown that combining TF-IDF with n-gram features will significantly improve performance relative to simple bag-of-words representations, demonstrating the strength of statistical text modeling in detecting misinformation (Figure 3).

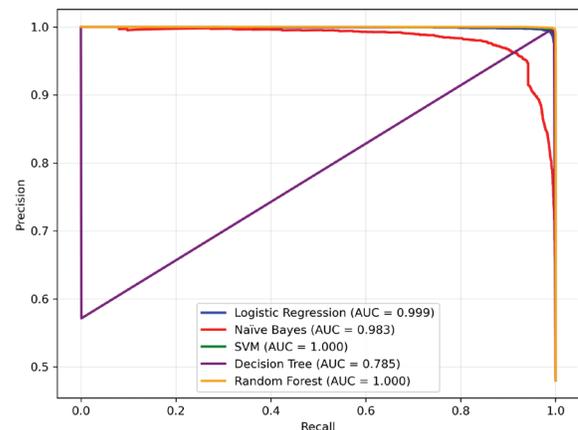


Figure 3. Precision and recall matrices in different models for the ISOT Dataset (44,898 articles). Random Forest has the highest area under the curve (AUC) of 1.

DISCUSSION

The expected findings reinforce the potential of classical machine learning techniques for scalable fake news detection. Despite the popularity of deep learning, this research demonstrates that simpler models, when paired with robust text representation, can achieve comparable accuracy while remaining interpretable and computationally efficient. Logistic Regression and SVM

models, in particular, offer consistent performance across datasets and minimal overfitting risk, making them ideal for real-time deployment (Figure 4).

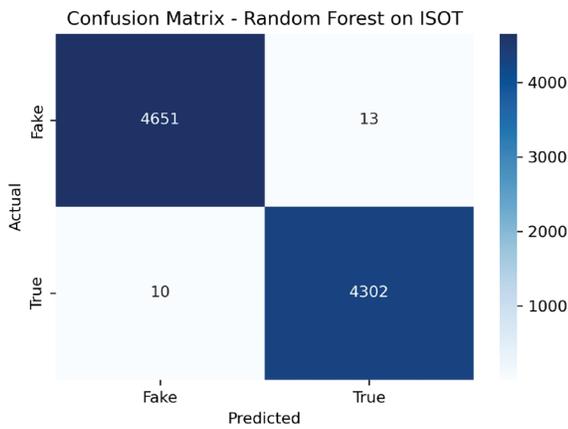


Figure 4. Confusion matrix of the best model (Random Forest) for the ISOT (44,898 articles), where it predicted an accuracy of 8953, and it only missed about 23, which demonstrates strong performance.

Interpretability remains a central advantage of classical approaches. Coefficients in Logistic Regression and feature importance scores in Random Forests can help identify key linguistic patterns, such as emotionally charged adjectives or sensationalist verbs, that differentiate fake from legitimate articles. These insights not only enhance transparency but also support media literacy and public understanding of the dynamics of misinformation (Figure 5).

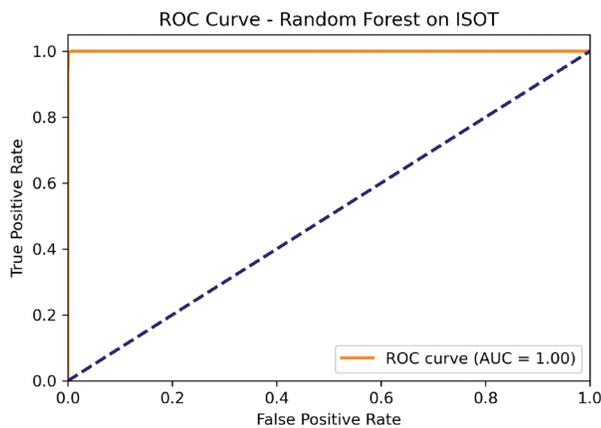


Figure 5. Area under the curve of the best model (Random Forest) for the ISOT Dataset (44,898 articles), where it has a true positive rate of around 100%.

Random Forests reduce variance by averaging over multiple decision boundaries, avoiding the overfitting prominent in decision trees alone. Transformer architectures, by contrast, excel by contextualizing words in both directions and capturing nonlocal dependencies, a key capability for distinguishing subtle indicators, sarcasm, or coordinated deception. Very high accuracy on ISOT but significant failure on LIAR points to deep model sensitivity to input structure and class balance. Models trained on full-length news fail to adapt to microtexts, exposing the need for hybrid systems and possibly domain-adaptive training (Figure 6).

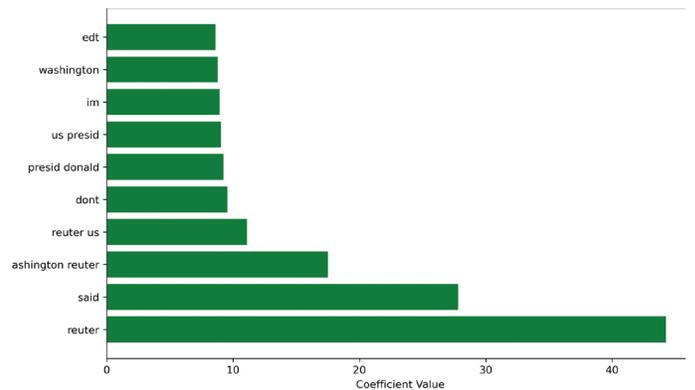


Figure 6. The top 10 feature importance of the Logistic Regression model during training. The model decides to base it on the most important feature through their coefficient values.

Stylistic cues, such as emotive language, repetition, or attribution, supplement factual assessment, illustrating the convergence between linguistic, psychological, and algorithmic analysis. Real-world deployment must contend with adversarial attacks, evolving misinformation tactics, domain drift, and the necessity for low-latency processing (Figure 7).

Nevertheless, limitations exist. Frequency-based models depend heavily on lexical patterns and may struggle with nuanced semantic deception or multilingual data. Future research should integrate semantic embeddings, topic modeling, or multimodal inputs (e.g., images, metadata) to enhance detection accuracy. Moreover, ethical considerations around dataset bias, fairness, and potential censorship must be addressed as these systems evolve. Automated systems must balance accuracy with the harms caused by false positives, respect for free expression, and regulatory requirements.

- 2018.8620051
5. Hiramath CK, Deshpande GC. Fake News Detection Using Deep Learning Techniques. In: 2019 1st International Conference on Advances in Information Technology (ICAIT). Chikmagalur, India: IEEE. 2019 [cited 2025 Nov 9]; p.411-415. <https://doi.org/10.1109/ICAIT47043.2019.8987258>
 6. Alshuwaier FA, Alsulaiman FA. Fake News Detection Using Machine Learning and Deep Learning Algorithms: A Comprehensive Review and Future Perspectives. *Computers*. 2025 Sept 16; 14 (9): 394. <https://doi.org/10.3390/computers14090394>
 7. Khan JY, Khondaker MdTI, Afroz S, Uddin G, Iqbal A. A benchmark study of machine learning models for online fake news detection. *Mach Learn Appl*. 2021 June; 4: 100032. <https://doi.org/10.1016/j.mlwa.2021.100032>
 8. Alghamdi J, Luo S, Lin Y. A comprehensive survey on machine learning approaches for fake news detection. *Multimed Tools Appl*. 2023 Nov 9; 83 (17): 51009-67. <https://doi.org/10.1007/s11042-023-17470-8>
 9. Sharma P, Nagpal T, Shrivastava G, Kumar JD. Machine Learning for Fake News Detection: A Survey. In: 2023 5th International Conference on Advances in Computing, Communication, Control and Networking (ICAC3N). Greater Noida, India: IEEE; 2023; p.850-5. <https://doi.org/10.1109/ICAC3N60023.2023.10541528>
 10. Teo TW, Chua HN, Jasser MB, Wong RTK. Integrating Large Language Models and Machine Learning for Fake News Detection. In: 2024 20th IEEE International Colloquium on Signal Processing & Its Applications (CSPA). Langkawi, Malaysia: IEEE; 2024; p.102-7. <https://doi.org/10.1109/CSPA60979.2024.10525308>
 11. Alghamdi J, Lin Y, Luo S. A Comparative Study of Machine Learning and Deep Learning Techniques for Fake News Detection. *Information*. 2022 Dec 12; 13 (12): 576. <https://doi.org/10.3390/info13120576>
 12. Gravanis G, Vakali A, Diamantaras K, Karadais P. Behind the cues: A benchmarking study for fake news detection. *Expert Syst Appl*. 2019 Aug; 128: 201-13. <https://doi.org/10.1016/j.eswa.2019.03.036>
 13. Varma R, Verma Y, Vijayvargiya P, Churi PP. A systematic survey on deep learning and machine learning approaches of fake news detection in the pre- and post-COVID-19 pandemic. *Int J Intell Comput Cybern*. 2021 Oct 4; 14 (4): 617-46. <https://doi.org/10.1108/IJICC-04-2021-0069>
 14. D'Souza KM, French AM. Fake news detection using machine learning: an adversarial collaboration approach. *Internet Res*. 2024 Sept 30; 34 (5): 1664-78. <https://doi.org/10.1108/INTR-03-2022-0176>
 15. Mishra S, Shukla P, Agarwal R. Analyzing Machine Learning Enabled Fake News Detection Techniques for Diversified Datasets. Hashmi MF, editor. *Wireless Commun Mob Comput*. 2022 Jan; 2022 (1): 1575365. <https://doi.org/10.1155/2022/1575365>
 16. Jiang T, Li JP, Haq AU, Saboor A, Ali A. A Novel Stacking Approach for Accurate Detection of Fake News. *IEEE Access*. 2021; 9: 22626-39. <https://doi.org/10.1109/ACCESS.2021.3056079>
 17. Jain A, Shakya A, Khatter H, Gupta AK. A Smart System for Fake News Detection Using Machine Learning. In: 2019 International Conference on Issues and Challenges in Intelligent Computing Techniques (ICICT). Ghaziabad, India: IEEE. 2019 [cited 2025 Nov 9]; p.1-4. <https://doi.org/10.1109/ICICT46931.2019.8977659>
 18. Sharma U, Saran S, Patil SM. Fake News Detection using Machine Learning Algorithms. *Int J Eng Res*. 2020; 9 (3).
 19. Baarir NF, Djeflal A. Fake News Detection Using Machine Learning. In: 2020 2nd International Workshop on Human-Centric Smart Environments for Health and Well-being (IHSH). Boumerdes, Algeria: IEEE; 2021; p.125-130. <https://doi.org/10.1109/IHSH51661.2021.9378748>
 20. Al-Tarawneh MAB, Al-irri O, Al-Maaitah KS, Kanj H, Aly WHF. Enhancing Fake News Detection with Word Embedding: A Machine Learning and Deep Learning Approach. *Computers*. 2024 Sept 19; 13 (9): 239. <https://doi.org/10.3390/computers13090239>