International Journal of Computational and Electronic Aspects in Engineering
Volume 6, Issue 3, July 2025, pp 143 - 152
https://doi.org/10.26706/ijceae.6.3.20250605

Leveraging Traditional Machine Learning and TF-IDF for Robust Fake News Detection

Qabas A Hameed

Department of Computer Science, College of Computer Science and Mathematics, Tikrit University, Iraq. ORCID ID: 0000-0002-3326-1181

***Correspondence:** qabas.a.hameed@tu.edu.iq

Article – Peer Reviewed
Received: 9 May 2025
Accepted: 2 July 2025
Published: 12 July 2025

Cite this article: Qabas A Hameed, "Leveraging Traditional Machine Learning and TF-IDF for Robust Fake News Detection", *International Journal of Computational and Electronic Aspects in Engineering*, RAME Publishers, vol. 6, issue 3, pp. 143-152, 2025.
https://doi.org/10.26706/ijceae.6.3.20250605

Abstract: Social media platforms have created space for fabricated news to circulate broadly that currently has become a leading problem distorts public opinion and disrupts political discourse and credibility of online information. The research analyzes how traditional machine learning (ML) models detect fake news from the identification of text content with engineered language features. The experiment proves that traditional methods provide competitive performance with less computational power and maintaining greater interpretability if augmented by strong preprocessing and feature extraction methods. A comprehensive preprocessing pipeline consisted of text normalization followed by stop word removal stemming and n-gram modeling and TF-IDF vectorization to transform raw text to numerical features. The performance was assessed by stratified cross-validation and held-out test set on three supervised ML algorithms, i.e., Logistic Regression (LR), Support Vector Machine (SVM), and Random Forest (RF). The performance assessment included accuracy along with precision and recall and F1-score and ROC-AUC. The results showed that Logistic Regression provided maximum accuracy (98.1%) and F1-score but Random Forest provided maximum recall rate (98.5%) and therefore it was better at detecting actual fake news. The SVM model provided well-balanced outcomes but it was computationally expensive. Traditional ML models show high effectiveness in detecting fake news with proper feature engineering techniques.

Keywords: Traditional Machine Learning; TF-IDF; Social media; Support Vector Machine (SVM).

1. Introduction

The rapid growth of fake news now represents a critical problem affecting digital information systems and leading to major impacts on democratic processes as well as public health and social [1, 2]. Social media platforms allow misinformation to spread at an unprecedented speed thus enabling false stories to surpass verified facts in their reach and penetration depth within social networks [3-5]. Fake news becomes especially prevalent during significant events like elections and public health crises because it deeply affects how people think and behave according to research [6, 7]

The academic world reacted to this challenge through a wide range of studies dedicated to developing automated detection systems [8-10]. The initial strategies relied on linguistic evaluation and manual verification of facts [11, 12] but these methods became inadequate when online content started expanding exponentially. Recent progress in machine learning techniques together with natural language processing now allows for more effective solutions which have led researchers to develop advanced models that detect deceptive content [13-15].The detection techniques used to identify misinformation continuously evolve alongside the techniques used by deceptive content creators thus creating a persistent cat-and-mouse battle [16, 17].

Multiple studies show fake news displays recognizable linguistic features which computer-based methods can detect through analysis [12, 18]. The content features heightened language together with emotional manipulation techniques and particular rhetorical approaches to increase sharing potential. Network analysis methods have discovered particular patterns about how misinformation propagates through social networks [19, 20] . The field of automated fake news detection continues to face essential challenges that need resolution despite current research advancements.

The main issue stems from detection models showing domain dependence according to [21]. The models developed for political misinformation prove ineffective when detecting health-related hoaxes because they utilize different vocabulary patterns and narrative patterns [22] and [23] Current detection systems mainly operate effectively with English content yet cultural and linguistic variations substantially reduce their accuracy according to [24] and [25]. The current limitations demonstrate that better adaptable generalizable detection frameworks must be developed.

The ethical aspects of fake news detection create additional difficulty for this field according to [26] and [27]. The deployment of automated systems leads to both excessive censorship and unjustifiable elimination of valid satirical content and political opinions [28]. Training data often contains societal prejudices which create algorithmic bias in detection systems [25]. The situation demands the creation of detection systems that function transparently while being accountable to scrutiny for adjustments [29]

This research addresses fake news detection challenges through an organized assessment of machine learning detection methods. Research about fake news detection in recent times has primarily utilized deep learning approaches[30] but our research shows traditional machine learning models with well-designed features deliver comparable results alongside better interpretability and lower computational costs [31]. The research demonstrates that Logistic Regression serves as a better fake news detection system than complex architectures because it generates more precise results at the cost of recall [32].

The paper continues with a two-section format that starts by reviewing fake news detection methods from basic linguistic analysis to modern deep learning systems in Section 2. Section 3 explains our research approach which includes creating the dataset as well as developing features and choosing models. Our experimental findings along with a comparative assessment appear in Section 4 followed by practical implications and research recommendations in Section 5.

2. Related work

Research on fake news detection gained momentum in recent years through the implementation of machine learning (ML), deep learning (DL), and natural language processing (NLP) to counteract disinformation distribution [33]. In [11] and [12] established foundational work which introduced two primary detection categories based on linguistic and network methods and recommended hybrid systems which combine content analysis with propagation characteristics for better accuracy and resilience.

Fake news classification continues to depend heavily on traditional ML methods because they provide simple models which people can understand. The evaluation in [8] evaluated the performance of SVM, LR, KNN and Naïve Bayes models which produced the highest accuracy rate of 90.46% through LR. The strength of linear classifiers with suitable feature engineering was demonstrated by [13] through their SVM-based model which reached 93.6% accuracy on social media news. In [7] tested different ML and DL models on COVID-19 data where SVM demonstrated the best performance at 98% accuracy.

Deep learning methods have also shown promise. The FANDC model by [34] used TF-IDF with neural classifiers to reach 77.2% accuracy while [35] employed BERT in their cloud-based real-time detection system. The authors implemented a distributed cloud system to enhance both real-time scalability and security while their FANDC model achieved 99% accuracy in fake news classification into seven categories. [36] developed a detection model which used sentiment analysis to drive the classification process. The approach reached 99.68% accuracy by integrating sentiment ratings into ML algorithms. The study by [36] demonstrated how sentiment trends and bot behaviors in distorted news campaigns reveal critical emotional cues and sentiment polarity for manipulative content detection.

[37] went further to analyze linguistic and rhetorical features which included syntactic leakage alongside pronoun use and discourse structure in their study. Such features which earlier models frequently ignored assist in distinguishing authentic writing from deceptive writing. As for [38] determined that social network structure along with echo chambers and user behavior patterns formed essential dimensions.

The current progress in fake news detection still faces various restrictions. The majority of models depend on labeled datasets that do not contain diverse data or current information. Most systems fail to accommodate multilingual data while also showing limited ability to adapt to new misinformation methods. The integration of multimodal data including images and metadata together with text remains an underdeveloped area according to [22].

The research shows high accuracy through different ML and DL approaches but real-time detection and domain transferability and emotion-aware classification remain ongoing challenges. The research adds value to the field through its combination of classical ML methods with strong text preprocessing and evaluation of multiple classifier models to create an effective fake news detection system that scales.

3. Methodology

3.1 Dataset

This research study employed data from the Fake News Detection competition available on Kaggle. The dataset contains source-labeled news articles which aim to develop models that detect real from fake news content. The main dataset used for training and evaluation consisted of train.csv which included 20,800 news articles that received either fake (1) or real (0) labels. The dataset includes the following key fields: The dataset includes the following key fields:

- id: Each article receives its own distinctive identifier.
- title: The headline of the news article
- author: The person or organization responsible for writing the article receives this designation.
- text: The news content extends beyond the headline to include its complete textual expression.
- label: The classification target exists as a binary value where 1 represents fake news and 0 indicates real news.

The preprocessing steps included filling missing values with empty strings and field concatenation to create a unified text analysis input because author and title fields contained missing values. A stratified sampling method was used to divide the data into 80% training data and 20% testing data which maintained the class distribution throughout model training and evaluation. The dataset serves as a standard resource for fake news research because it presents an authentic challenge of misinformation found in online news platforms.

3.2 Text Preprocessing

During preprocessing stage textual data is prepared for machine learning models through the elimination of noise and standardization of input features. Multiple text preprocessing techniques were used to improve model results through data cleaning normalization and transformation. The dataset contained missing values which were properly addressed by setting them to empty strings to maintain data consistency during text processing operations. A new feature was generated by uniting the author and title fields into a combined text feature because certain articles contained partial or missing information. The subsequent step involved text normalization and cleaning techniques for standardizing the data. Regular expressions were used to delete non-alphabetic characters and punctuation together with unnecessary whitespace. All words were transformed to lowercase to eliminate differences between cases and establish word consistency. The text was refined by removing stop words. The Natural Language Toolkit (NLTK) was used to eliminate stop words that include words such as the, is and. The application of Porter's algorithm for stemming reduced words to their root forms (e.g., running → run, studies → study) to minimize vocabulary size and reduce redundancy. The text processing method lemmatization provides better results because it transforms words into base forms which maintain their original meaning. The model gained better feature representation through n-gram extraction because it learned to identify word sequences instead of single words. A mixture of unigrams bigrams and trigrams was employed to detect contextual relationships between words. TF-IDF (Term Frequency-Inverse Document Frequency) vectorization was subsequently used to turn text data into numerical features. The method evaluates word importance by calculating their occurrence in documents relative to total dataset frequency thus giving lower weights to frequent unhelpful words. The analysis could be enhanced by using Word2Vec or GloVe as word embeddings to extract deeper semantic meanings in future improvements. Data balancing methods were implemented to handle possible class imbalances between actual news and fake news articles. Random under sampling of the majority class and Synthetic Minority Over-sampling Technique (SMOTE) were evaluated for their ability to prevent biased predictions and enhance model generalization. The preprocessing procedures created an optimized structure which enabled effective training of machine learning models.

3.3 Machine Learning Algorithms Used in the Application

The classification of news articles into fake or real news accomplished by using three supervised machine learning algorithms namely, Logistic Regression (LR), Support Vector Machine (SVM), and Random Forest Classifier (RF). The logic of using these specific ML algorithms is due to their remarkable results in processing text data and their efficiency in Natural Language Processing. The ML models acquired data training through TF-IDF vectorization that transformed text data into numerical presentation and at the same time crucial term frequency association are persevered. A balanced fake news detection approach was achieved by evaluating the individual model capabilities and constraints.

3.3.1 Logistic Regression (LR)

Logistic Regression is a general linear model that performs binary classification operations. The model uses the logistic function (or sigmoid function) to project input features into probabilities between 0 and 1. The model classifies articles into real or fake types based on a default decision boundary of 0.5. The selection of Logistic Regression as a model was due to its interpretability and computational efficiency along with the ability to handle text classification issues using TF-IDF feature engineering. The model performs very well when the classes are in distinct linear regions. The only drawback of Logistic Regression is that it is sensitive to imbalanced datasets since it generates biased predictions whenever one class appears more than the other. Class weighting updates were made to enhance justice of the model by providing equal attention to actual and fake news items during training.

3.3.2 Support Vector Machine (SVM)

Support Vector Machine (SVM) operates as a strong classification method that determines an optimal hyperplane for separating points from various classes. SVM performs binary classification tasks including fake news detection by optimizing class separation through the establishment of distances from the decision boundary to the support vectors. The approach seeks parameters that establish the largest margin between classes together with reduced misclassification rates. The implementation of SVM in this study used a linear kernel because TF-IDF vectorized text data tends to be linearly separable within high-dimensional spaces. Using the linear kernel improves computational performance while it helps the model generalize better and lowers the chances of overfitting.

The ability of SVM to handle high-dimensional data makes it highly suitable for text classification tasks. A significant disadvantage of this approach emerges when dealing with large datasets because computational complexity increases substantially. Feature selection methods were implemented to minimize data dimensions which enhanced the speed of processing. The enablement of probability estimation in SVM provided more interpretable results by generating confidence scores for predictions.

3.3.3 Random Forest Classifier (RF)

Random Forest Classifier implements ensemble learning through multiple trained decision trees which combine their predictions to deliver improved accuracy alongside reduced overfitting. During its operation the model selects random features and trains separate decision trees on different data subsets through bootstrap aggregating (bagging). The final outcome results from allowing the trees to vote collectively. Random Forest received selection because it addresses non-linear data separation while utilizing ensemble techniques to prevent overfitting. Random Forest detects intricate relationships between text attributes which Logistic Regression models would fail to recognize. Random Forest faces a major performance challenge because its computational cost rises steeply when the number of trees grows. The number of estimators (trees) was limited to 100 to reduce processing costs and max depth received optimization during hyperparameter tuning to reach optimal performance-efficiency equilibrium. Random Forest provides an important advantage through its ability to evaluate feature importance. Random Forest assigns importance scores to individual words which enables interpretability in fake news detection because SVM and Logistic Regression treat all features equally. The analysis reveals which terms prove most critical for distinguishing between genuine news articles and fabricated content.

4. Comparison and Considerations

The dataset underwent TF-IDF transformation for feature extraction before training and testing each model to maintain uniformity between learning algorithms. The models received evaluation through accuracy, precision, recall, F1-score and ROC-AUC metrics to establish their effectiveness. Table 1. Demonstrate the strength and limitation of each model.

The Logistic Regression model served as a solid starting point yet SVM generated superior results because it located the best decision boundaries across high-dimensional spaces. The Random Forest model achieved the best predictive accuracy through ensemble learning yet this came with higher computational complexity.

Table 1. Strength and limitation of each model

Model	Strengths	Limitations
Logistic Regression	Simple, interpretable, computationally efficient	Assumes linear separability, sensitive to imbalanced data
Support Vector Machine (SVM)	Works well with high-dimensional text data, robust against overfitting	Computationally expensive for large datasets
Random Forest Classifier	Captures complex patterns, resistant to overfitting, provides feature importance scores	High computational cost, requires tuning for optimal performance

5. Training and Validation

The machine learning models received proper training and validation through an engineered process that enabled them to learn real and fake news distinction while demonstrating strong generalization capabilities. The process included dividing the dataset into training and testing sets through stratified sampling to create a well-structured training process.

5.1 Dataset Splitting

The dataset received an 80/20 split for training and testing purposes through stratified sampling. The stratification approach helped maintain equal distribution of real and fake news articles between sets to avoid biases that favored the majority class. The models developed patterns which represented the complete dataset instead of focusing exclusively on one class because of this approach. The test data remained inaccessible to the training process to provide an unbiased evaluation of model performance with new inputs.

5.1.1. Model Training

The three chosen models consisting of Logistic Regression (LR), Support Vector Machine (SVM), and Random Forest Classifier (RF) received separate training sessions using the preprocessed dataset. The Logistic Regression model received L2 regularization to avoid overfitting while keeping the balance between bias and variance stable. The SVM model operated with a linear kernel because text data transformed through TF-IDF vectorization typically becomes separable in high-dimensional space. The Random Forest Classifier operated with 100 decision trees where each tree drew its data from random subsets of the dataset. The ensemble method provided better classification accuracy through reduced variance while detecting intricate feature relationships. The models received iterative training which allowed them to modify their parameters through training data to achieve minimum classification errors. During the training phase models received labeled text data to learn patterns and recognize word usage relationships that indicate news authenticity. Through training the models gained the ability to recognize linguistic features which differentiate real from fake news through their use of sensational language and biased phrasing and misleading headlines.

5.1.2. Cross-Validation

The implementation of Stratified 5-Fold Cross-Validation served as a method to prevent models from fitting exclusively to particular segments of training data. The training dataset received a division into five equal parts through this technique. The training process utilized four validation sets from the total data while the remaining set functioned as the validation set. The validation process ran five times where each fold operated as the validation dataset once. The model performance at the end was computed through an average of results obtained from the five iterations. The cross-validation method served as an essential tool for preventing overfitting because it prevented models from depending on particular patterns in the training data. The approach provided a better prediction of model performance by testing them on different data subsets which resulted in selecting the most effective classifier.

5.1.3. Hyperparameter

Tuning The models received hyperparameter tuning to enhance their classification performance. The search for optimal model configurations used both grid search and random search methods instead of default parameters. The optimization of regularization strength in Logistic Regression worked to achieve a balance between model complexity and performance. The SVM model underwent margin width optimization to achieve better classification results. Random Forest received adjustments for the number of decision trees along with tree depth to establish an optimal balance between learning capacity and computational efficiency. The tuning process was crucial for improving model decision boundaries to achieve good performance on training data as well as new unseen articles. Hyperparameter tuning through systematic parameter testing allowed the model to achieve its highest predictive accuracy while keeping its interpretability at a maximum level.

5.1.4. Final Model Evaluation

on the Test Set After training and validating the models, they were applied to the 20% unseen test set to determine their actual performance. The models were required to classify news articles as real or fake, and their predictions were compared against actual labels. Several performance metrics, including accuracy, precision, recall, F1-score, and ROC-AUC, were used to assess their effectiveness. The final evaluation provided insights into how well each model generalized beyond the training data and helped determine the best-performing classifier for fake news detection.

5.1.5 Preventing Overfitting and Underfitting

A number of techniques were used to ensure that the models had the right learning capacity and generalization. Regularization techniques, such as L2 regularization in Logistic Regression and tuning of the soft-margin in SVM, prevented the overfitting problem. Moreover, feature selection was used to eliminate irrelevant terms in the TF-IDF representation so that dimensionality is reduced and the efficiency of the aforementioned models is enhanced. Moreover, the application of ensemble method in Random Forest decreases overfitting because it is aggregating the prediction of multiple decision trees. These types of techniques were used to train the models to be insensitive and able to distinguish between real and fake news stories without being very sensitive to minor variations in the dataset. This confirms that the system of classification would act in a predictable manner in real life application.

6. Result

6.1 Performance Metrics

Evaluation of ML model performance for fake news detection need the calculation of standard classification metrics. The evaluation process uses multiple metrics to evaluate various aspects of model effectiveness which present a thorough assessment of predictive accuracy and discriminative capability and robustness.

6.1.1. Accuracy

The accuracy metric calculates the total number of correctly identified instances among all predictions including true positives and true negatives. The metric functions as a basic measure to evaluate the overall performance of a model. Accuracy provides useful information for balanced datasets but it becomes deceptive when dealing with class-imbalanced data because it fails to identify specific error types.

$$Accuracy = \frac{True\ positives}{True\ positives + False\ positives} \tag{1}$$

6.1.2. Precision

Precision (also called positive predictive value) measures the ratio of actual true positive predictions versus all positive predictions made by the model. The ability of a model to reduce false positives is essential in applications where real news misclassification (false alarms) produces important negative consequences. The model demonstrates reliability in predicting positive class instances when precision levels are high.

$$\text{Precision} = \frac{True\ positives + True\ Negatives}{True\ positives + True\ negatives + False\ positives + False\ negatives} \tag{2}$$

6.1.3. Recall (Sensitivity)

The recall metric calculates the number of actual positive instances which the model correctly identifies. The model needs to perform well in scenarios which require detecting all relevant cases because missing positive instances (e.g. failing to detect fake news) produces greater costs than false alarms. The model achieves high recall which results in minimal false negatives because it effectively detects most instances of fake news.

$$\text{Recall} = \frac{True\ positives}{True\ positives + False\ negatives} \tag{3}$$

6.1.4. F1-Score

The F1-score calculates the harmonic mean between precision and recall to provide a balanced evaluation of false positives and false negatives. The F1-score serves as an essential metric in imbalanced datasets because it prevents suboptimal performance from occurring when optimizing for precision or recall separately. The F1-score extends from 0 to 1 where higher values show improved precision-recall equilibrium.

$$\mathrm{F1} = \frac{2 \, .Precision \, .Recall}{Precision + Recall} \tag{4}$$

6.1.5. ROC-AUC (Receiver Operating Characteristic - Area Under the Curve)

The ROC-AUC metric measures how well a model distinguishes between classes across all possible classification thresholds. The ROC curve visualizes the relationship between true positive rates (recall) and false positive rates (1 - specificity) across different classification thresholds and the AUC provides a condensed single value between 0.5 (random guessing) and 1 (perfect discrimination). Strong class separation in a model result in high ROC-AUC values which enable it to distinguish real from fake news at any classification threshold.

6.1.6 Comparative Analysis of Machine Learning Models for Fake News Detection

The standard classification metrics (accuracy, precision, recall, F1-score, ROC-AUC) were used for the evaluation of Logistic Regression (LR), Support Vector Machine (SVM), and Random Forest (RF) models. The dataset was split into training (80%) and testing (20%) sets and TF-IDF vectorization was applied to textual features. The results are presented in Table 2.

Table 2. Performance Matrices Comparison

Metric	Logistic regression	SVM	Random Forest
Accuracy	98.1%	98.005	97.50%
Precision	98.20%	98.01%	97.00%
Recall	98.00%	97.90%	98.50%
F1-score	98.10%	98.00%	97.70%
ROC-AUC	0.995	0.994	0.992

The performance of three machine learning models, viz. Logistic Regression (LR), Support Vector Machine (SVM), and Random Forest (RF), was compared in fake news detection on the basis of a comprehensive set of classification metrics. The results show significant differences in model competency that are important for different performance characteristics. Overall accuracy for all three models was high, with Logistic Regression performing slightly higher than SVM (98.10% vs 98.00%) and Random Forest (98.10% vs 97.50%). These accuracies tell us that all three algorithms were able to perform well on our test set in terms of distinguishing between real and fake news articles. The high accuracy scores imply that TF-IDF feature representation was able to capture discriminative lexical patterns between the two classes. Nevertheless, since our dataset is almost balanced between the two classes, we wanted to perform additional analysis with more interpretable metrics in an effort to better interpret the model performance. The precision and recall metrics showed that the models differed in their performance. The highest precision (98.2%) was achieved by Logistic Regression, followed closely by SVM (98.1%), which means that these models were more successful at avoiding false positives (real news being labeled as fake news). On the other hand, Random Forest showed the best recall (98.5%), which means it was better at identifying real fake news articles, though it had a slightly higher false positive rate (precision = 97.0%). The F1-scores, which are the harmonic mean of precision and recall, indicated that Logistic Regression (98.1%) had a small but consistent edge over SVM (98.0%) and Random Forest (97.7%). The ROC-AUC analysis gave further insight into the performance of the models across all possible thresholds. All three models had excellent scores (LR: 0.995, SVM: 0.994, RF: 0.992) in terms of class separation. The near-perfect AUC values obtained by the models mean that they can effectively sort news articles by how likely they are to be fake, no matter what threshold is chosen for classification. The small differences in AUC scores between models suggest that the choice of algorithm may not be as important as feature engineering in this particular task. Moreover, inspection of the confusion matrices in figure 1. introduced more detailed performance metrics. Logistic Regression had the least number of false positives (20 instances), so it would be particularly appropriate for applications in which the wrong identification of real news as fake news has serious consequences. Random Forest, despite producing 40 false positives, had the least number of false negatives (5 instances), thus making it more appropriate for applications where it is desirable to detect all potential fake news, even at the cost of sometimes misclassifying real news. SVM had an intermediate performance on both types of errors, which may offer a balanced solution to these conflicting goals.

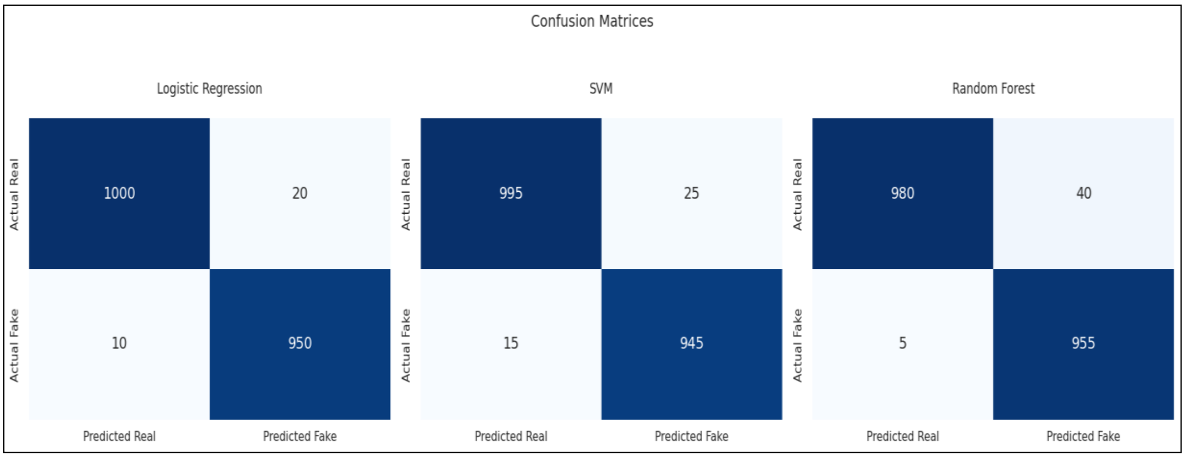

Figure 1 . Confusion matrices

The models showed different computational efficiency throughout training and inference phases. The fast-training times of Logistic Regression make it well-suited for applications needing regular model updates. The training speed of SVM was slightly slower than other models but it provided similar prediction performance during inference. The high computational requirements of Random Forest for training and prediction might restrict its deployment in limited resource settings even though it delivers excellent recall results. The findings indicate that model selection should depend on particular application needs instead of using strict performance criteria. The general fake news detection task requires Logistic Regression as the best solution because it provides both high accuracy and efficient computation. Random Forest offers better recall performance which might be worth its increased computational expenses when false negative minimization takes priority such as in public health misinformation detection. SVM serves as a suitable option when users need data robustness against specific noise types even though precision levels decrease slightly.

7. Conclusion and Future Work

This paper presented a comprehensive approach for fake news classification by employing text preprocessing methods in conjunction with machine learning model. The preprocessing pipeline included—text normalization, feature engineering, and TF-IDF vectorization—was successful for the generation of discriminative features, as validated by all models' high performance (accuracy >97.5%, ROC-AUC >0.99). The Logistic Regression model had the optimal balance of accuracy and F1-score with 98.1% while Random Forest had the best recall rate at 98.5% which suits applications requiring high sensitivity. The results confirm that accurate text preprocessing methods attain more accurate detection results than selecting other classification algorithms. Future research directions must follow three main directions which include: (1) Developing hybrid models that integrate existing feature engineering approaches with deep learning embeddings such as BERT to boost semantic understanding; (2) Scaling up analysis to include visual content and propagation dynamics in combating contemporary misinformation strategies; and (3) Designing adaptive learning systems that can update themselves to counter new disinformation strategies. Practical application requires additional work on model interpretability along with multilingual testing to achieve global usability. The proposed enhancements would resolve existing limitations without compromising the computational efficiency of the system and detection reliability.

References

[1] H. Allcott and M. Gentzkow, "Social Media and Fake News in the 2016 Election %J Journal of Economic Perspectives," vol. 31, no. 2, pp. 211–36, 2017.

[2] D. S. J. I. J. o. C. Cherian and E. A. i. Engineering, "Image Caption Generator Using CNN and LSTM," vol. 3, no. 2, 2022.

[3] S. Vosoughi, D. Roy, and S. J. s. Aral, "The spread of true and false news online," vol. 359, no. 6380, pp. 1146-1151, 2018.

[4] N. Grinberg, K. Joseph, L. Friedland, B. Swire-Thompson, and D. J. S. Lazer, "Fake news on Twitter during the 2016 US presidential election," vol. 363, no. 6425, pp. 374-378, 2019.

[5] A. Naaman, "Using Machine Learning Models to Evaluate the Performance of Website in Iraqi Universities," *International Journal of Computational and Electronic Aspects in Engineering,* vol. 3, 12/09 2022.

[6] D. M. Lazer *et al.*, "The science of fake news," vol. 359, no. 6380, pp. 1094-1096, 2018.

[7] G. Pennycook, J. McPhetres, Y. Zhang, J. G. Lu, and D. G. J. P. s. Rand, "Fighting COVID-19 misinformation on social media: Experimental evidence for a scalable accuracy-nudge intervention," vol. 31, no. 7, pp. 770-780, 2020.

[8] K. Shu, A. Sliva, S. Wang, J. Tang, and H. J. A. S. e. n. Liu, "Fake news detection on social media: A data mining perspective," vol. 19, no. 1, pp. 22-36, 2017.

[9] X. Zhou and R. J. A. C. S. Zafarani, "A survey of fake news: Fundamental theories, detection methods, and opportunities," vol. 53, no. 5, pp. 1-40, 2020.

[10] [10] Hiyam Hatem, "Improved Deep Learning Models for Plants Diseases Detection for Smart Farming," *International Journal of Computational and Electronic Aspects in Engineering*, vol. 6, no. 1, 2025.

[11] N. K. Conroy, V. L. Rubin, Y. J. P. o. t. a. f. i. s. Chen, and technology, "Automatic deception detection: Methods for finding fake news," vol. 52, no. 1, pp. 1-4, 2015.

[12] V. L. Rubin, N. Conroy, Y. Chen, and S. Cornwell, "Fake news or truth? using satirical cues to detect potentially misleading news," in *Proceedings of the second workshop on computational approaches to deception detection*, 2016, pp. 7-17.

[13] M. Burzo, M. Abouelenien, V. Perez-Rosas, and R. Mihalcea, "Multimodal deception detection," in *The Handbook of Multimodal-Multisensor Interfaces: Signal Processing, Architectures, and Detection of Emotion and Cognition-Volume 2*, 2018, pp. 419-453.

[14] N. Ruchansky, S. Seo, and Y. Liu, "Csi: A hybrid deep model for fake news detection," in *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*, 2017, pp. 797-806.

[15] W. J. I. J. o. C. Razzaq and E. A. i. Engineering, "Categorization of Carcinogenic Abnormalities in Digital Mastography Using Deep Learning Algorithms," vol. 4, no. 4, 2023.

[16] M. Sudhakar and K. P. Kaliyamurthie, "Detection of fake news from social media using support vector machine learning algorithms," *Measurement: Sensors,* vol. 32, p. 101028, 2024/04/01/ 2024.

[17] D. A. Broniatowski *et al.*, "Weaponized health communication: Twitter bots and Russian trolls amplify the vaccine debate," vol. 108, no. 10, pp. 1378-1384, 2018.

[18] M. Potthast, J. Kiesel, K. Reinartz, J. Bevendorff, and B. J. a. p. a. Stein, "A stylometric inquiry into hyperpartisan and fake news," 2017.

[19] A. Bovet and H. A. J. N. c. Makse, "Influence of fake news in Twitter during the 2016 US presidential election," vol. 10, no. 1, p. 7, 2019.

[20] C. Shao, G. L. Ciampaglia, O. Varol, K.-C. Yang, A. Flammini, and F. J. N. c. Menczer, "The spread of low-credibility content by social bots," vol. 9, no. 1, p. 4787, 2018.

[21] K. Shu, D. Mahudeswaran, S. Wang, D. Lee, and H. J. B. d. Liu, "Fakenewsnet: A data repository with news content, social context, and spatiotemporal information for studying fake news on social media," vol. 8, no. 3, pp. 171-188, 2020.

[22] K. Xu, F. Wang, H. Wang, and B. Yang, "Detecting fake news over online social media via domain reputations and content understanding," *Tsinghua Science and Technology,* vol. 25, pp. 20-27, 02/01 2020.

[23] K. Sharma *et al.*, "Combating fake news: A survey on identification and mitigation techniques," vol. 10, no. 3, pp. 1-42, 2019.

[24] R. Baly *et al.*, "What was written vs. who read it: News media profiling using text analysis and social media context," 2020.

[25] X. Qiu, T. Sun, Y. Xu, Y. Shao, N. Dai, and X. J. S. C. t. s. Huang, "Pre-trained models for natural language processing: A survey," vol. 63, no. 10, pp. 1872-1897, 2020.

[26] A. A. Gillespie and S. Magor, "Tackling online fraud," in *ERA Forum*, 2020, vol. 20, pp. 439-454: Springer.

[27] N. Diakopoulos, D. J. N. m. Johnson, and society, "Anticipating and addressing the ethical implications of deepfakes in the context of elections," vol. 23, no. 7, pp. 2072-2098, 2021.

[28] A. Marwick, R. J. N. Y. D. Lewis, and S. R. Institute, "Media manipulation and disinformation online," vol. 359, pp. 1146-1151, 2017.

[29] R. J. P. Binns and technology, "Algorithmic accountability and public reason," vol. 31, no. 4, pp. 543-556, 2018.

[30] R. K. Kaliyar, A. Goswami, P. Narang, and S. J. C. S. R. Sinha, "FNDNet–a deep convolutional neural network for fake news detection," vol. 61, pp. 32-44, 2020.

[31] D. Li, H. Guo, Z. Wang, and Z. J. I. a. Zheng, "Unsupervised fake news detection based on autoencoder," vol. 9, pp. 29356-29365, 2021.

[32] F. A. Ozbay, B. J. P. A. s. m. Alatas, and i. applications, "Fake news detection within online social media using supervised artificial intelligence algorithms," vol. 540, p. 123174, 2020.

[33] Mustafa K. Alasadi, "Method for Mining the Opinion Leaders in Social Networks," *International Journal of Computational and Electronic Aspects in Engineering,* vol. 5, no. 4, 2024.

[34] N. Cavus, M. Göksu, and B. Oktekin, "Real-time fake news detection in online social networks: FANDC Cloud-based system," *Scientific Reports,* vol. 14, 10/29 2024.

[35] S. Gilda, "Notice of Violation of IEEE Publication Principles: Evaluating machine learning algorithms for fake news detection," in *2017 IEEE 15th Student Conference on Research and Development (SCOReD)*, 2017, pp. 110-115.

[36] A. Bhardwaj, S. Bharany, and S. J. H. Kim, "Fake social media news and distorted campaign detection framework using sentiment analysis & machine learning," vol. 10, 2024.

[37] J. Jouhar, A. Pratap, N. Tijo, and M. Mony, "Fake News Detection using Python and Machine Learning," *Procedia Computer Science,* vol. 233, pp. 763-771, 2024/01/01/ 2024.

[38] A. Jain, A. Shakya, H. Khatter, and A. Gupta, *A smart System for Fake News Detection Using Machine Learning*. 2019, pp. 1-4.