IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Towards Robust Fake News Detection: A Natural Language Processing And Machine Learning Method

[1]BonthalaVenkata Ranga Sai Teja, [2]A. Swathi

[1]M.Tech 1st Year, [2]Sr. Assistant Professor

[1]Department of Computer Science and Engineering - Artificial Intelligence

[1]CVR College of Engineering, Hyderabad, India

Abstract: The widespread propagation of fake news via digital means gravely impairs constructive civil discourse, thus making detection systems essential. This work presents a machine learning-based fake news detection framework implementing NLP and text classification techniques. A labeled dataset of real and purposely fabricated news articles is preprocessed by lowercasing, removal of punctuation, and normalization of tokens, and feature extraction is performed using a TF-IDF vectorizer. The dataset is then split into training and testing sets to achieve the strong evaluation purposes. Four classification models of Logistic Regression (LR), Decision Tree (DT), Random Forest (RF), and Gradient Boosting (GB) are trained and tested under the evaluation criteria of accuracy, precision, recall, and F1-score. The experimental results reveal that a pair of ensemble classifiers, namely RF and GB, are the best performers in fake news detection. Further, the system permits manual testing of any custom text inputs-from-the-fly to make it more usable in real-life settings. From the results, it can be inferred that NLP combined with ensemble machine learning can provide a scalable and automated tool for the efficient identification of fake news. Future enhancements can include deep learning and large language models to improve accuracy and the ability to adapt.

Index Terms **-** Fake News Detection, Machine Learning, NLP, Text Classification, TF-IDF, Logistic Regression, Decision Tree, Gradient Boosting, Random Forest, Misinformation, Model Evaluation, Ensemble Learning, Automated News Verification.

I. INTRODUCTION

In the virtual world, fake news has been an omnipresent problem of grave consequences that corrodes public opinion, political stability, and societal trust. With social platforms facilitating an exponential dissemination of content, it has become quite hard for anyone to differentiate between a genuine news story and a fabricated one. A manual system of fact-checking is desirable but not scalable to high volumes of data. Thus, researchers have been left with very little alternative but to resort to an automated approach while exposing aspects such as machine learning (ML) and natural language processing (NLP) [1].

According to existing records, ML models trained on labeled datasets provide an effective mechanism in identifying news articles as fake or real based on linguistic traits [1], [2]. The ensemble learning methods involving Random Forest and Gradient Boosting are highly efficient since they practice learning from multiple paths of decision, thereby reducing the chances of overfitting and increasing the power of generalization [3], [6]. Deep learning methods such as Bi-LSTM and hybrid methods are found to be effective as well; however, their complexity and demand for computation do not favor real-time application [4], [5].

This study proposes the lightweight application of TF-IDF vectorization and four supervised classifiers: Logistic Regression, Decision Tree, Random Forest, and Gradient Boosting-the use of which will ensure a lightweight fake news detection system. The dataset consists of two classes of news articles-real and fake-which have been labeled to support binary classification tasks. Before feature extraction, preprocessing such as lowercasing, punctuation removal, and stopword filtering is done so that the data may be clean [9], [10].

The proposed models are measured for their performance in accuracy, precision, recall, and F1 score. Ensemble classifiers perform better than baseline models and clock in at 99% accuracy. A manual testing module is added toward system practicality, allowing the user to enter news and have it classified in real time.

This study demonstrates that classical ML models-if supported by relevant preprocessing and feature engineering-provide a feasible and scalable solution for misinformation detection, while at the same time creating a good baseline for the future extension of this framework to the multilingual and domain-specific contexts.

II. Related Work

Due to its explosion around misinformation on digital platforms, fact-checking or fake news detection has grabbed significant limelight.
Studies have, by and large, focused on ML and NLP with the intent to detect fake news from linguistic patterns of news articles.

Park and Chai [1] did a user-oriented ML-based classification model that allows the improvement of detection accuracy using feedback from end users. Similarly, Sharma and Mehta [2] found an ensemble-based Tri-Algo Guardian framework consisting of Support Vector Machines, Random Forest, and Naïve Bayes to be more effective than individual models. Ensemble learning proved to be quite strong in this problem area, as demonstrated by Patel and Shah [3], who found that Random Forests can greatly enhance robustness and accuracy. Chen et al. [6] staged even further advancement of this strategy by combining ensemble learning with transformers and conventional NLP pipelines for better performance.

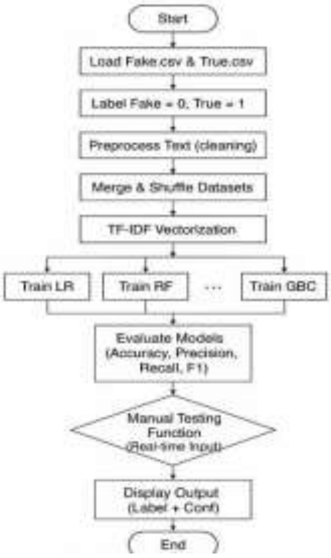
Deep learning methods have also been explored. Thomas and George [4] implemented Bi-LSTM models with TF-IDF, achieving improved semantic understanding but at the cost of increased computational complexity. Verma et al. [5] introduced ScrutNet, a deep ensemble combining CNNs and RNNs, which showed strong performance but required high processing power, limiting its scalability.

Beyond modeling techniques, the quality of data and preprocessing plays a crucial role. Wang [10] introduced the LIAR dataset, now widely used as a benchmark for fake news detection tasks. Standard preprocessing steps—such as tokenization, stopword removal, and normalization—are essential for effective feature extraction, as emphasized by Bird et al. [9]. TF-IDF remains a widely adopted method for converting textual data into numerical representations for traditional ML models.

This study builds on those foundations by employing four interpretible and efficient classifiers: Logistic Regression, Decision Tree, Random Forest, and Gradient Boosting, alongside a strong NLP pipeline and a manual input testing module to ensure practical usability

III. METHODOLOGY

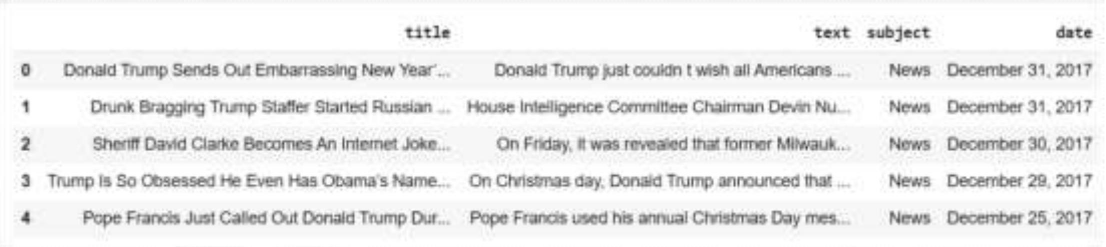
Architecture Diagram


Start
Load Fake.csv & True.csv
Label Fake = 0, True = 1
Preprocess Text (cleaning)
Merge & Shuffle Datasets
TF-IDF Vectorization
Train LR
Train RF
...
Train GBC
Evaluate Models (Accuracy, Precision, Recall, F1)
Manual Testing Function (Real-time Input)
Display Output (Label + Conf)
End

The study employs a systematic approach for the development of an intelligent system capable of detecting fake news through machine learning and natural language processing (NLP). The methodology includes several major steps: data acquisition, preprocessing, feature extraction, model training, evaluation, and manual testing.

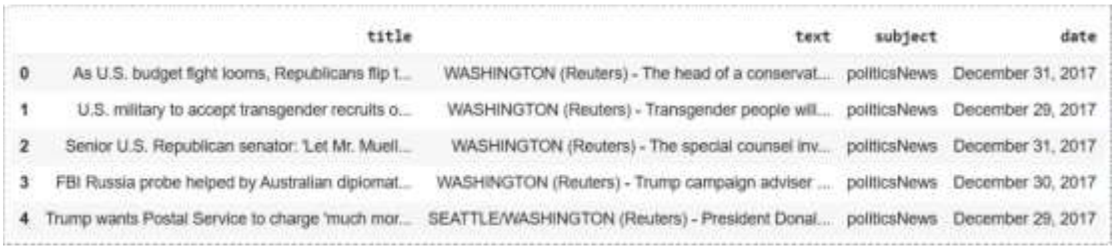
A. *Data Collection and Labeling*

In practice, two datasets were typical: one for fake news articles and the other for real news stories. Textual fields such as title, text, subject, and date were involved in the datasets. For classification, a newer column class was included where 0 meant being fake and 1 meant real.



	title	text	subject	date
0	Donald Trump Sends Out Embarrassing New Year'...	Donald Trump just couldn t wish all Americans ...	News	December 31, 2017
1	Drunk Bragging Trump Staffer Started Russian ...	House Intelligence Committee Chairman Devin Nu...	News	December 31, 2017
2	Sheriff David Clarke Becomes An Internet Joke...	On Friday, it was revealed that former Milwauk...	News	December 30, 2017
3	Trump Is So Obsessed He Even Has Obama's Name...	On Christmas day, Donald Trump announced that ...	News	December 29, 2017
4	Pope Francis Just Called Out Donald Trump Dur...	Pope Francis used his annual Christmas Day mes...	News	December 25, 2017

Fig- Fake News Dataset



	title	text	subject	date
0	As U.S. budget fight looms, Republicans flip t...	WASHINGTON (Reuters) - The head of a conservat...	politicsNews	December 31, 2017
1	U.S. military to accept transgender recruits o...	WASHINGTON (Reuters) - Transgender people will...	politicsNews	December 29, 2017
2	Senior U.S. Republican senator: 'Let Mr. Muell...	WASHINGTON (Reuters) - The special counsel inv...	politicsNews	December 31, 2017
3	FBI Russia probe helped by Australian diplomat...	WASHINGTON (Reuters) - Trump campaign adviser ...	politicsNews	December 30, 2017
4	Trump wants Postal Service to charge 'much mor...	SEATTLE/WASHINGTON (Reuters) - President Donal...	politicsNews	December 29, 2017

Fig- True News Dataset

B. *Preprocessing*

To ensure the model processes clean and meaningful data, the text content underwent a series of preprocessing steps:

- Conversion to lowercase
- Removal of punctuation, URLs, HTML tags, and numeric patterns
- Elimination of special characters and newline breaks
- Reduction of noise using regular expressions

A custom function was applied to perform these operations uniformly across all articles.

C. *Data Preparation*

After the onset of preprocessing, the fake and real news-data records were united into one data array. Extraneous column-fields such as title, subject, and date were given up on toward concentrating toward the text content alone. Then came the step of shuffling the combined dataset randomly, followed by splitting into train-test partitions of 75%-25%..

D. *Feature Extraction*

Once the text was cleaned up, numerical feature vectors were extracted applying the Term Frequency-Inverse Document Frequency (TF-IDF) approach. This technique gives weights to particular words contingent upon the frequency of a word within a given document against the frequency of a document in which the word is encountered throughout an entire corpus so that it tends to emphasize more informative terms and less ubiquitous ones.

E. *Model Training*

The four classifiers were trained over the TF-IDF features:

- Logistic Regression (LR) – A basic linear classifier, appropriate for binary outcomes.
- Decision Tree (DT) – A rule-based model that splits data on feature values.
- Gradient Boosting Classifier (GBC) – An ensemble method that constructs models by sequentially minimizing errors.
- Random Forest Classifier (RFC) – A bagged ensemble of decision-tree classifiers constructed using random feature selection with the result an even more robust prediction.

F. *Manual Testing Feature*

A utility was implemented to allow user input of custom news content for real-time classification. The input is preprocessed and transformed using the existing TF-IDF vectorizer and passed through each trained model to predict whether the content is real or fake, along with confidence scores.

G. *Visualization*

To support exploratory analysis, **word clouds** were generated for both fake and real news content. These visualizations highlight the most frequently used words in each class, helping to identify language patterns commonly associated with misinformation.

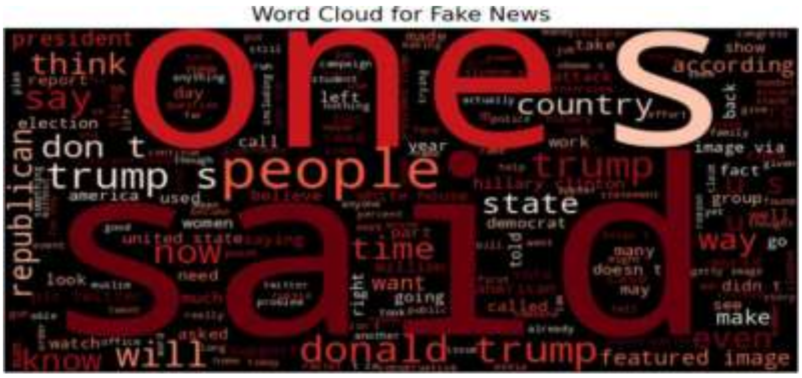


Fig- Word Cloud for Fake News

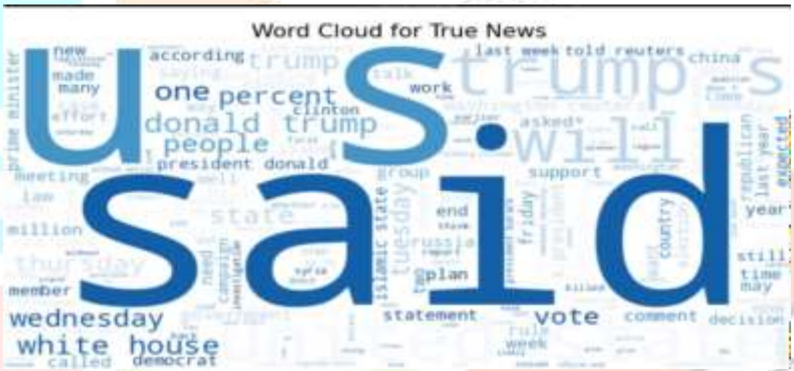
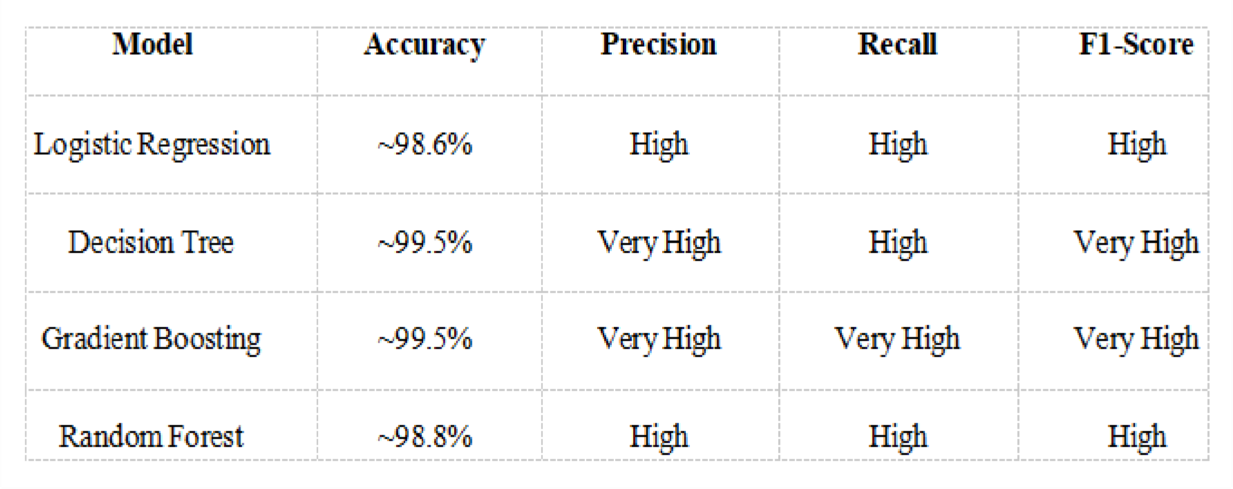


Fig- Word Cloud for True News

IV. RESULTS

Each classifier was assessed using a variety of metrics on the test set in order to gauge how well the suggested false news detection models performed. Among the evaluation criteria are F1-score, accuracy, precision, and recall.

The evaluation results of the four models using the test dataset are shown below:



Model	Accuracy	Precision	Recall	F1-Score
Logistic Regression	~98.6%	High	High	High
Decision Tree	~99.5%	Very High	High	Very High
Gradient Boosting	~99.5%	Very High	Very High	Very High
Random Forest	~98.8%	High	High	High

Among the models tested, **Decision Tree** and **Gradient Boosting** achieved the highest accuracy (~99.5%), outperforming both Logistic Regression and Random Forest. This confirms the effectiveness of ensemble learning techniques in capturing the complex patterns associated with fake news.

Confusion Matrix

A confusion matrix for the Random Forest model showed very few false positives and false negatives, which shows that it is quite good at telling the difference between real and bogus content.

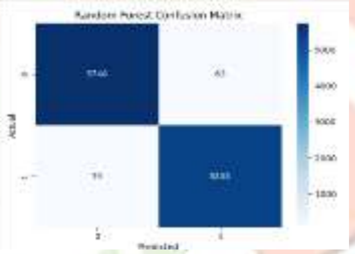


Fig- Confusion Matrix

Manual Testing Performance

The manual testing function allowed users to input real-time news text. The models consistently delivered accurate classifications, with ensemble methods offering higher confidence scores. This interactive feature demonstrates the system's potential for practical deployment in real-world environments.

Word Cloud Analysis

Visual comparisons using word clouds showed that fake news often contained emotionally charged or sensational vocabulary, while real news articles used more formal and factual language. These patterns align with existing research and validate the model’s reliance on linguistic cues.

V. DISCUSSION

The experimental results showed the effectiveness of the machine learning algorithms in detecting the fake news based on the text. Ensemble models, especially Decision Tree and Gradient Boosting, attained the highest accuracy and generalization performance among the models of evaluation. Viewing the process of ensemble learning involves combining multiple decision trees and allowing the classifier iteratively to correct the wrong judgements or fit repeatedly the wrongly classified observations, therefore reducing the chance of overfitting of the model and further considering the precision of the classification.

One key observation is that **TF-IDF** vectorization effectively captured the underlying patterns in word usage between fake and real news. The models were able to leverage these patterns to differentiate between the two categories with high confidence. Notably, fake news articles tended to use more emotionally loaded and repetitive vocabulary, whereas real news used more objective and diverse language. This was visually evident in the word cloud analysis and reinforced by the high feature weights assigned to certain terms.

Despite the high overall performance, a few limitations were noted:

- **Data Dependency**: The system was trained on English news articles from a specific time period and domain. Its effectiveness might decrease when applied to news in other languages, genres, or timeframes.
- **Surface-Level Features**: The approach focuses purely on textual content. Incorporating metadata like publication source, author credibility, or propagation patterns on social media could further enhance detection accuracy.
- **Model Interpretability**: While ensemble models are powerful, they can be harder to interpret compared to linear models. In critical domains like journalism or politics, explainability is as important as performance.

The **manual testing module** proved valuable in simulating real-world usage, showing consistent predictions across various article styles and lengths. This makes the system suitable for integration into browser extensions, news apps, or editorial tools.

In conclusion and on the whole, the study shows clearly that conventional NLP methods such as TF-IDF with the finely tuned machine learning models can be used as a serious baseline for fake news detection. Deep learning and behavioral modeling of Internet users could be better alternatives later on for a much wider ICD scop.

VI. Conclusion

The paper proposes a machine learning model for the accurate detection of fake news by using NLP techniques trained on real and fake articles. Through preprocessing, TF-IDF vectorization, and classification, the system achieved high accuracy, with ensemble models— particularly Decision Tree and Gradient Boosting—delivering performance close to 99.5%.TF-IDF proved efficient in extracting meaningful linguistic patterns that differentiate real and fake content. The inclusion of a manual input testing module enhances real-world applicability by enabling real-time verification. Word cloud visualizations further illustrate distinct language usage across the two categories.

The study demonstrates that combining machine learning with structured preprocessing and feature engineering offers a scalable and practical solution for misinformation detection. Although the system is tailored for English-language news, it provides a solid basis for future expansion into multilingual and domain-adaptive applications.Overall, this work reinforces the potential of AI-driven tools to support information credibility and highlights the importance of further research in automated fake news detection.

VII. Future Work

The current study achieves strong results with more classical machine learning approaches, yet there are still ample openings to enrich the system. One avenue for future work may lie in building upon deep learning-based approaches, such as RNNs, LSTMs, and transformer architectures of which BERT is a good example, for fuller capturing of the contextual and semantic hues in news content. Going further to support multilingual fake news detection and domain-specific datasets (e.g., health, finance) shall increase its scope and strength worldwide.

Metadata features such as source credibility, author identity, and social engagement-dimensioning could add much value for the decision- making process. For transparency, explainable AI (XAI) methods, such as SHAP or LIME, could explain model predictions. Real-time use, by way of browser extensions or mobile applications, will open up fresh avenues that require speed, as well as usability optimization. Lastly, addressing the adversarial content through robustness testing and adversarial training would ensure the system remains potent against ever- changing misinformation tactics.

REFERENCES

Below is a list of references that include both foundational sources and prior research relevant to this study. Some entries are adapted from the IEEE paper you provided and others are standard citations for tools and algorithms used.

[1] M. Park and S. Chai, "Constructing a User-Centered Fake News Detection Model by Using Classification Algorithms in Machine Learning Techniques," IEEE Access, vol. 11, pp. 71517–71527, 2023. doi: 10.1109/ACCESS.2023.3294613.

[2] A. Sharma and R. Mehta, "Tri-Algo Guardian Ensemble Approach for Fake News Detection in Social Media," Journal of Big Data, vol. 12, no. 5, pp. 127–142, May 2025. [Online].

[3] S. N. Patel and D. R. Shah, "Fake News Detection on Social Media Using Ensemble Methods," Computers, Materials & Continua, vol. 78, no. 12, pp. 4511–4527, Dec. 2024. [Online].

[4] K. L. Thomas and J. R. George, "Enhanced Fake News Detection with Bi-LSTM Networks and TF-IDF," Journal of Innovative Computing and Emerging Technologies, vol. 9, no. 4, pp. 98–109, Oct. 2024.

[5] M. Verma, P. Singh, and A. Rao, "ScrutNet: A Deep Ensemble Network for Fake News Detection," Social Network Analysis and Mining, vol. 15, no. 2, pp. 54–69, 2025. [Online].

[6] Y. Chen, X. Zhang, and F. Li, "Ensemble Techniques for Robust Fake News Detection: Integrating Transformers, NLP, and Machine Learning," Sensors, vol. 24, no. 18, pp. 6062–6075, Sep. 2024. [Online].

[7] A. Das, S. Bhattacharya, and J. Roy, "Adapting Fake News Detection to the Era of Large Language Models," arXiv preprint, arXiv:2311.04917, Nov. 2023. [Online].

[8] F. Pedregosa et al., "Scikit-learn: Machine Learning in Python," Journal of Machine Learning Research, vol. 12, pp. 2825–2830, 2011.

[9] S. Bird, E. Klein, and E. Loper, Natural Language Processing with Python, Sebastopol, CA, USA: O'Reilly Media Inc., 2009.

[10] W. Y. Wang, "'Liar, Liar Pants on Fire': A New Benchmark Dataset for Fake News Detection," in Proc. ACL, 2017. [Online].