

A Computational Investigation of Fabricated News Identification Using TF-IDF and Machine Learning

Asane Shubhangi Bhimraj¹ and Dr. H. B. Jadhav²

Student, Department of Computer Engineering¹

HOD, Department of Computer Engineering²

Adsul's Technical Campus, Chas, Ahilyanagar, India

Abstract: *The rapid expansion of digital communication has made it increasingly difficult to distinguish reliable journalism from fabricated or deliberately misleading information. False stories, altered claims, and unsupported statements can circulate within seconds through websites, social-media platforms, and online discussion forums. Manually checking every article is time-consuming and difficult to scale. To address this challenge, this study develops a text-based fake-news detection system that combines Natural Language Processing, Term Frequency-Inverse Document Frequency (TF-IDF), and supervised Machine Learning.*

The news dataset was cleaned, normalised, and converted into TF-IDF feature vectors before model training. Logistic Regression, Multinomial Naive Bayes, and Random Forest were trained and evaluated under the same experimental conditions. After preprocessing, 38,605 valid records remained, and the TF-IDF representation was restricted to 5,000 important features. Model performance was measured using accuracy, precision, recall, F1-score, and confusion matrices. Random Forest achieved the highest accuracy of 99.62%, followed by Logistic Regression at 98.54% and Multinomial Naive Bayes at 93.46%. These findings demonstrate that traditional machine-learning methods can effectively identify patterns associated with fabricated news within the selected textual dataset. However, the results remain dependent on the dataset, preprocessing strategy, feature configuration, and model settings used in the experiment.

Keywords: Fake News Detection, Natural Language Processing, TF-IDF, Machine Learning, Logistic Regression, Multinomial Naive Bayes, Random Forest, Text Classification.

I. INTRODUCTION

The Internet has transformed the way news is produced, distributed, and consumed. Readers can now access information almost instantly through online newspapers, social-media platforms, blogs, and messaging applications. Although this accessibility has improved the speed of communication, it has also created favourable conditions for misleading and fabricated information to spread rapidly. A false story can reach a large audience before its claims have been properly examined, making it difficult for readers to determine whether the information is trustworthy.

The problem is not limited to completely invented stories. Misleading headlines, selective reporting, exaggerated statements, and information presented without proper context can also influence public understanding. As digital content continues to grow, relying only on manual verification becomes increasingly impractical. Human fact-checking remains valuable because it can consider context and intent, but it requires considerable time, specialised knowledge, and effort.

Machine Learning provides a practical way to assist this process. By studying patterns in previously labelled examples, a classification model can learn to identify textual characteristics associated with different news categories. In a text-



based system, an article is first cleaned and transformed into numerical features. These features are then supplied to a trained model, which estimates whether a new article resembles fabricated or authentic news.

This study investigates a complete text-classification workflow beginning with dataset preparation and text cleaning. The processed content is represented using TF-IDF and evaluated with three supervised algorithms: Logistic Regression, Multinomial Naive Bayes, and Random Forest. All models use the same dataset, preprocessing steps, feature representation, and train-test division. This controlled setting makes the comparison more meaningful because differences in performance can be associated primarily with the classification algorithms.

The main objective is to examine how effectively these three models identify fabricated news from textual content. Accuracy, precision, recall, F1-score, and confusion matrices are considered so that the evaluation reflects more than a single performance value. The study also recognises that a text-only system cannot capture every aspect of misinformation. Future versions could incorporate publisher information, multilingual content, images, videos, social context, and other forms of supporting evidence.

II. EXISTING SYSTEM

Traditional news verification depends largely on human judgement. A person reads an article, identifies its important claims, and compares those claims with information obtained from credible sources. This approach can provide valuable context and help reveal subtle forms of manipulation. However, manually reviewing a very large collection of articles requires substantial time and sustained attention.

Automated text-analysis systems offer an alternative for handling large volumes of content. In a typical workflow, news records are collected, incomplete or irrelevant entries are removed, and the remaining text is converted into a numerical representation. A Machine Learning classifier can then learn from labelled examples and use the learned patterns to categorise previously unseen articles.

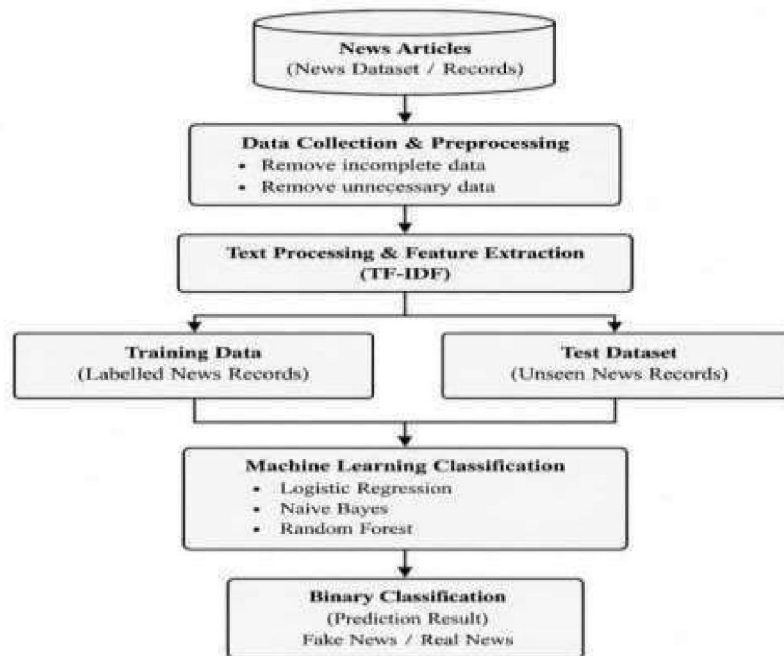
TF-IDF is frequently used for this purpose because it highlights terms that are important within particular documents while reducing the influence of words that occur throughout the entire collection. Once the text has been transformed into feature vectors, algorithms such as Logistic Regression, Naive Bayes, and Random Forest can be trained for binary classification.

Nevertheless, a text-based approach has clear limitations. The wording of an article may not reveal whether its publisher is reliable, whether an image has been taken out of context, or whether the information is supported by independent evidence. Language differences, writing style, duplicated content, and changes in public events can also influence model performance. Consequently, a high score on one dataset should not automatically be interpreted as universal reliability.

To maintain a fair comparison, this study applies the same dataset preparation, text-cleaning procedure, TF-IDF configuration, and train-test split to all three models. Only the classification algorithm is changed. This design allows the measured differences to be examined under consistent experimental conditions.



II. EXISTING SYSTEM



Explanation of Existing System Diagram:

The existing system shows how news articles are automatically classified as Fake News or Real News using TF-IDF and Machine Learning algorithms.

News Articles – News records are collected from the dataset.

Data Collection & Preprocessing – Incomplete, duplicate, or unnecessary information is removed to improve data quality. Text Processing & TF-IDF – The news text is converted into numerical features using TF-IDF (Term Frequency– Inverse Document Frequency).

Training and Test Data – The prepared dataset is divided into labelled training data and unseen test data.

Machine Learning Classification – Three algorithms are applied:

Logistic Regression

Naive Bayes

Random Forest

Binary Classification – The trained model predicts whether a news article is Fake News or Real News.

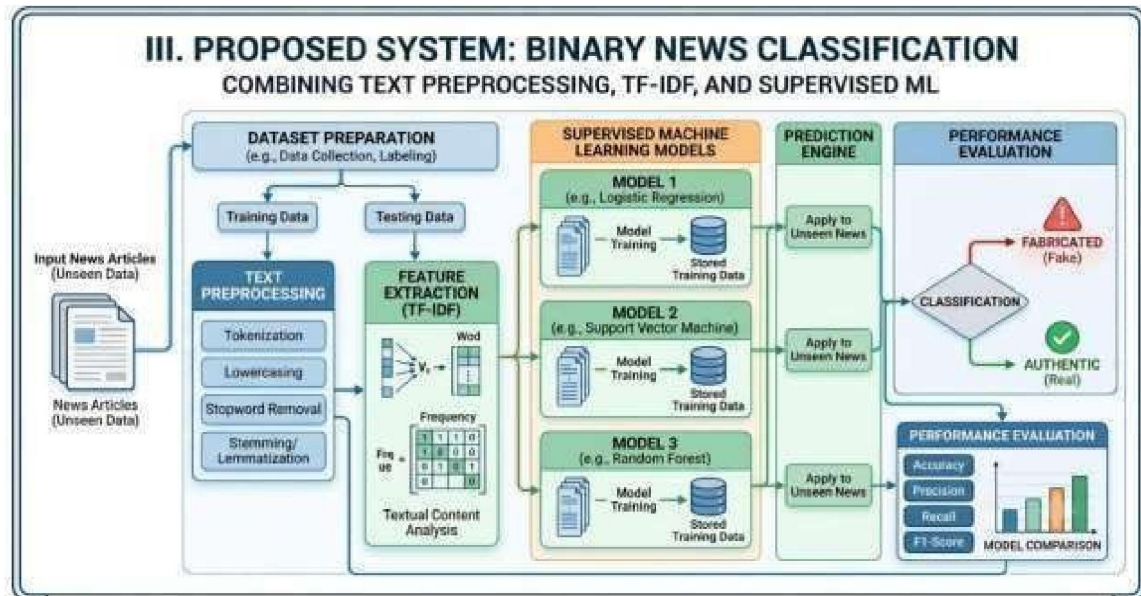
III. PROPOSED SYSTEM

The proposed system is a binary text-classification framework designed to distinguish fabricated news from authentic news. It combines data preparation, text preprocessing, TF-IDF feature extraction, supervised model training, prediction, and performance evaluation in a single workflow.

The process begins with a labelled news dataset. Unusable records are removed, and the remaining text is normalised to create a consistent input format. The cleaned content is then converted into numerical TF-IDF vectors. These vectors are used to train Logistic Regression, Multinomial Naive Bayes, and Random Forest models.

After training, the models receive unseen test records and predict their categories. The predictions are compared with the actual labels using several evaluation measures. Because every model receives the same processed input, the resulting comparison provides a clearer view of their relative behaviour.





A. Dataset

The dataset contains two labelled categories: fabricated news and authentic news. Each record includes a headline and related article content. The two categories are combined into a single dataset for supervised classification. Records that lack meaningful text or contain unusable information are excluded during preparation.

After preprocessing, the final dataset contains 38,605 valid records. Of these, 17,908 belong to the fabricated-news category, and 21,197 belong to the authentic-news category. The data is divided into training and testing subsets using an 80:20 ratio. Stratified splitting is applied so that both subsets retain a comparable proportion of the two classes.

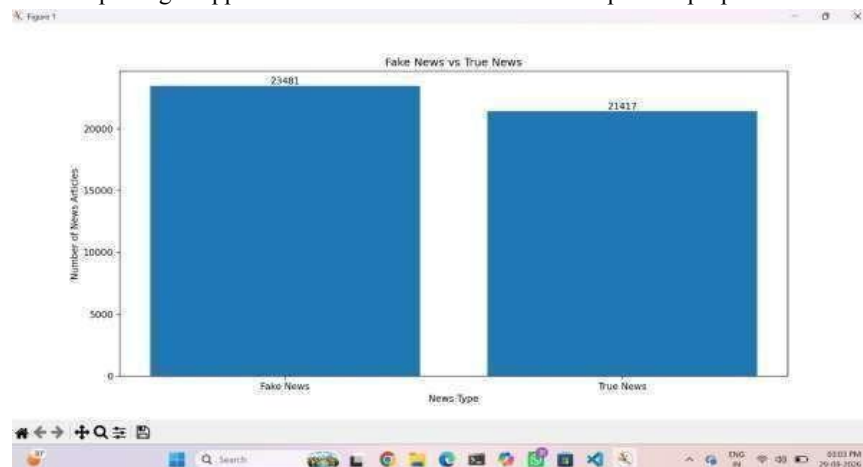


Fig. 1. Distribution of Fabricated and Authentic News Records

B. Text Preprocessing

Raw news content may contain missing values, inconsistent capitalisation, punctuation, special symbols, and other elements that do not contribute meaningfully to classification. The preprocessing stage removes records without useful content and converts the remaining entries into a consistent string format.



All text is converted to lowercase, and unnecessary characters are removed. This reduces superficial variation between documents. For example, the same word written with different capitalisation is treated as a single term rather than as multiple unrelated features. The resulting text is cleaner and more suitable for feature extraction.

C. TF-IDF Feature Extraction

Term Frequency–Inverse Document Frequency converts written language into numerical values that Machine Learning algorithms can process. The method assigns greater importance to terms that are frequent in a particular document but comparatively uncommon across the complete collection.

Terms that appear in almost every document generally receive lower weights because they provide limited information for distinguishing one article from another. In contrast, terms that occur more selectively may contribute more strongly to the classification process.

In this study, the TF-IDF vocabulary is limited to 5,000 features. The vectoriser is fitted only on the training data and subsequently applied to the test data. This procedure helps prevent information from the test set from influencing the construction of the feature space.

D. Machine Learning Models

The study evaluates three classification algorithms.

a) Logistic Regression :

Logistic Regression estimates the probability that an article belongs to one of two categories. Using the TF-IDF features, it learns a decision boundary that separates textual patterns associated with fabricated news from those associated with authentic news. Its relatively simple structure makes it efficient and suitable for high-dimensional text data.

b) Multinomial Naive Bayes

Multinomial Naive Bayes is a probability-based method widely used for document and text classification. It estimates how strongly different terms are associated with each news category and combines these probabilities when classifying a new article. Although the model uses a simplifying independence assumption, it can perform effectively on sparse word-based representations.

c) Random Forest

Random Forest is an ensemble algorithm that combines the predictions of multiple decision trees. Each tree examines the data from a slightly different perspective, and their combined output is used to produce the final classification. This ensemble structure can help the model capture varied relationships within the feature space and reduce dependence on any single decision tree.

E. Prediction and Evaluation

Once training is complete, each model is evaluated using the reserved test set. The predicted categories are compared with the known labels to calculate accuracy, precision, recall, and F1-score.

Confusion matrices are also generated to show the number of correct and incorrect predictions for each category. Considering these measures together provides a more balanced interpretation of model behaviour. Accuracy reflects overall correctness, precision indicates the reliability of positive predictions, recall measures the proportion of relevant records identified, and the F1-score balances precision with recall.



IV. EXPERIMENTAL RESULTS AND ANALYSIS

All three classifiers are evaluated using the same test records and the same TF-IDF representation. From the 38,605 valid records, 30,884 are used for model development, and 7,721 are reserved for testing according to the 80:20 division.

The evaluation considers accuracy, precision, recall, F1-score, and confusion matrices. This combination of measures makes it possible to examine both overall performance and the types of errors produced by each classifier.

A. Model Performance

Classifier	Accuracy	Precision	Recall	F1-Score
Logistic Regression	98.54%	98.18%	99.17%	98.67%
Multinomial Naive Bayes	93.46%	93.32%	94.88%	94.09%
Random Forest	99.62%	99.48%	99.83%	99.66%

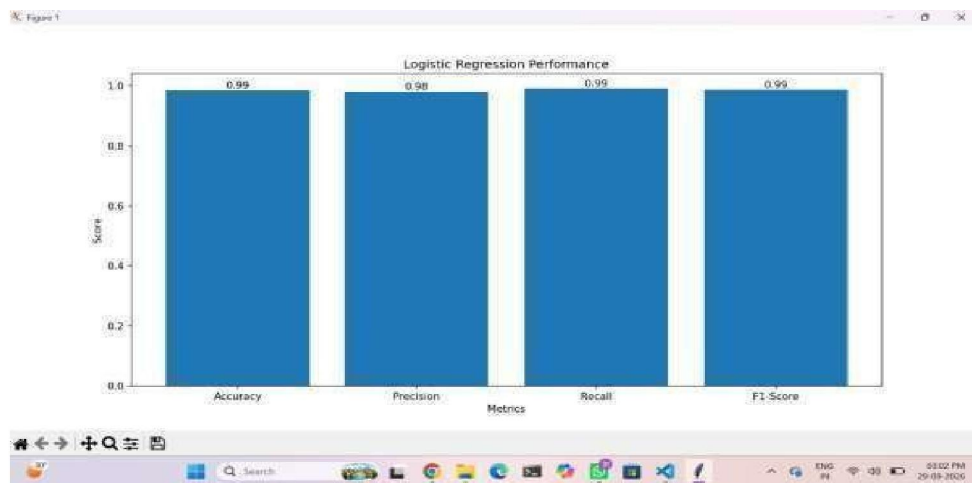


Fig. 2. Performance Metrics of Logistic Regression

The Logistic Regression model achieved an accuracy of 98.54% and an F1-score of 98.67% under the given experimental conditions.



Fig. 3. Performance Metrics of Multinomial Naive Bayes



The Multinomial Naive Bayes model achieved an accuracy of 93.46% and an F1-score of 94.09%.



Fig. 4. Performance Metrics of Random Forest

The Random Forest model achieved an accuracy of 99.62% and an F1-score of 99.66%.

The results show that the three models produced different performance values when tested under the same experimental conditions. These results are specific to the selected dataset, preprocessing steps, TF-IDF configuration, train-test division, and model settings. Therefore, they should not be considered universal performance values for other datasets.

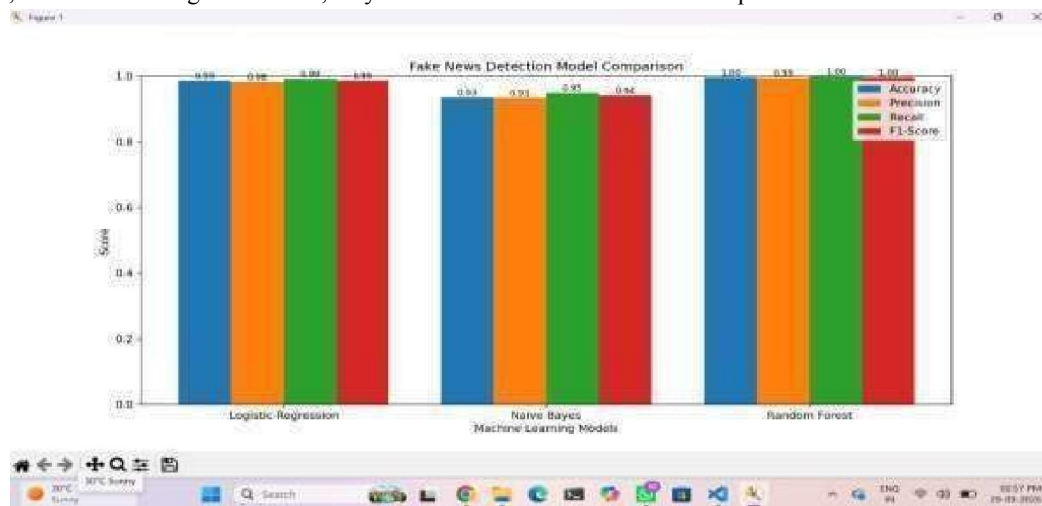


Fig. 5. Comparison of Classification Models

B. Confusion Matrix Analysis

A confusion matrix provides a detailed view of the predictions by showing correctly classified and misclassified records for each category.

For Logistic Regression, the confusion matrix is:

[[3404, 78],

[35, 4204]]





Fig. 6. Confusion Matrix of Logistic Regression

For Multinomial Naive Bayes, the confusion matrix is:

[[3194, 288],
[217, 4022]]



Fig. 7. Confusion Matrix of Multinomial Naive Bayes

For Random Forest, the confusion matrix is:

[[3460, 22],
[7, 4232]]





Fig 8. Confusion Matrix of Random Forest

The matrices show that the models made different numbers of classification errors. Random Forest produced the fewest misclassifications among the three models, which is consistent with its higher accuracy and F1-score. Multinomial Naive Bayes produced comparatively more incorrect predictions, indicating that its simplifying probability assumptions may not capture all the relationships present in the selected TF-IDF feature space.

C. Result Discussion

The experiment demonstrates that the classification algorithm influences the effectiveness of a TF-IDF-based fake-news detection system. Under the conditions used in this study, Random Forest achieved the highest accuracy at 99.62%, followed by Logistic Regression at 98.54% and Multinomial Naive Bayes at 93.46%.

Because all three models used the same dataset preparation, preprocessing procedure, feature representation, and train-test division, the comparison provides a consistent basis for evaluating their measured performance. Random Forest's stronger result suggests that its ensemble of decision trees was better able to separate the textual patterns present in this dataset.

Accuracy describes the overall percentage of correctly classified records. Precision indicates how often positive predictions are correct, while recall shows how many relevant records are successfully identified. The F1-score combines precision and recall into one measure. Confusion matrices complement these scores by revealing the specific distribution of correct and incorrect predictions.

The findings should not be generalised beyond the experimental setting without additional testing. Different datasets, languages, news sources, preprocessing procedures, feature limits, and model parameters may produce different outcomes. In practical applications, automated predictions should therefore assist human verification rather than replace critical evaluation entirely.

V. CONCLUSION

This study developed and evaluated a text-based fake-news detection system using Natural Language Processing, TF-IDF feature extraction, and supervised Machine Learning. The news records were cleaned, normalised, and converted into numerical feature vectors before being supplied to three classifiers: Logistic Regression, Multinomial Naive Bayes, and Random Forest.

The models were assessed using accuracy, precision, recall, F1-score, and confusion matrices. For the selected dataset and configuration, Random Forest achieved the highest accuracy of 99.62%, followed by Logistic Regression at 98.54% and Multinomial Naive Bayes at 93.46%.



The results indicate that TF-IDF combined with conventional Machine Learning can effectively classify textual news within the dataset examined. The study also highlights that model performance is closely connected to the quality and composition of the data, the preprocessing strategy, the feature representation, and the selected algorithm. Consequently, the reported results should be understood as evidence from a specific experiment rather than as a universal measure of fake-news detection capability.

Future research could extend the system by incorporating multilingual news, publisher credibility, article metadata, social-media behaviour, images, videos, and external fact-checking information. A multimodal approach may provide a more complete understanding of misinformation because it would consider not only the words used in an article but also its source, context, presentation, and supporting evidence.

VI. ACKNOWLEDGMENT

The authors sincerely thank the Department of Computer Engineering, Adsul's Technical Campus, Chas, Ahilyanagar, for providing the academic environment, facilities, and encouragement needed to complete this study. The authors are also grateful to the faculty members and contributors whose guidance, suggestions, and constructive feedback supported the development and evaluation of the proposed system.

REFERENCES

- [1] J. Alghamdi, S. Luo, and Y. Lin, "A comprehensive survey on machine learning approaches for fake news detection," *Multimedia Tools and Applications*, vol. 83, pp. 51009–51067, 2024.
- [2] "Fake News Detection Using Natural Language Processing and Machine Learning Techniques," in *Proc. 7th International Conference on Image Information Processing (ICIIP)*, Solan, India, 2023, doi:10.1109/ICIIP61524.2023.10537640.
- [3] A. Saeed and E. A. Solami, "Fake News Detection Using Machine Learning and Deep Learning Methods," *Computers, Materials & Continua*, vol. 77, no. 2, pp. 2079–2096, 2023, doi: 10.32604/cmc.2023.030551.
- [4] K. Szabó Nagy, J. Kapusta, and M. Munk, "Feature extraction from unstructured texts as a combination of the morphological and the syntactic analysis and its usage in fake news classification tasks," *Neural Computing and Applications*, vol. 35, pp. 22055–22067, 2023.
- [5] E. Essa, K. Omar, and A. Alqahtani, "Fake news detection based on a hybrid BERT and LightGBM models," *Complex & Intelligent Systems*, vol. 9, pp. 6581–6592, 2023, doi: 10.1007/s40747-023-01098-0.
- [6] K. S. S. Datta, G. Narasimha Naidu, S. Abhishek, and T. Anjali, "Enhancing Veracity: Empirical Evaluation of Fake News Detection Techniques," *Procedia Computer Science*, vol. 233, pp. 97–107, 2024, doi:10.1016/j.procs.2024.03.199.
- [7] Z. Tahat, A. Gharaibeh, M. Z. Tahat, W. B. Glisson, H. Alamleh, and X. Liu, "Social Media Fake News Detection Using Machine Learning Models and Feature Extraction Techniques," in *Proc. 2024 IEEE 15th Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON)*, pp. 134–139, 2024, doi:10.1109/UEMCON62879.2024.10754749.
- [8] M. A. Taha, H. D. A. Jabar, and W. K. Mohammed, "Fake News Detection Model Based on Machine Learning Algorithms," *Baghdad Science Journal*, vol. 21, no. 8, p. 2771, 2024, doi: 10.21123/bsj.2024.8710.
- [9] D. K. Nugroho, M. N. Fauzy, and K. R. Hidayat, "Evaluating Machine Learning Algorithms for Detecting Online Text-based Fake News Content," *International Journal of Computer and Information System*, vol. 6, no. 3, 2025, doi: 10.29040/ijcis.v6i3.253.
- [10] R. Pavithra Devi, K. Shesathri, Y. Madhumitha, U. Dinesh, and A. Mohanty, "A Scalable Approach to Fake News Detection Using TF-IDF and Logistic Regression," in *Proc. 2025 International Conference on Sensors and Related Networks (SENNET)*, pp. 1–6, 2025, doi: 10.1109/SENNET64220.2025.11135996.

