



The Scientific Temper

SEPTEMBER 2026 Vol. 17 (9): 6912-6927

ISSN: 0976-8653, E-ISSN: 2231-6396

DOI: 10.58414/SCIENTIFICTEMPER.2026.17.9.2520

A WEB OF SCIENCE JOURNAL

RESEARCH ARTICLE

A Novel Weighted Soft Voting Method for Fake News Detection in Social Media

S. VENKATESAN, DR. T. S. BASKARAN

Research Scholar, PG and Research Department of Computer Science, A.V.V.M Sri Pushpam College (Autonomous), Poondi, Thanjavur(Dt) (Affiliated to Bharathidasan University, Tiruchirappalli), TamilNadu, India. svenkat.subramani@gmail.com

Research Advisor and Associate Professor, PG and Research Department of Computer Science, A.V.V.M Sri Pushpam College (Autonomous), Poondi, Thanjavur(Dt) (Affiliated to Bharathidasan University, Tiruchirappalli), TamilNadu, India. t_s_baskaran@yahoo.com

Corresponding author: svenkat.subramani@gmail.com

ABSTRACT

Social media platforms have brought about a massive change in the way information is created, shared and used. This accessibility, which is widespread, has also accelerated the dissemination of fake news and created public confusion and the presence of misinformation that has consequences on public life. Lexical features are the main ones used in conventional machine learning methods and are limited in their ability to detect features of the news content because they do not have the ability to identify the complex semantic and contextual properties. In this regard, this study suggests a Hybrid Ensemble Learning (ML) Algorithm for Fake News identification Using Semantic and Contextual Features to solve some of these issues. The suggested model combines complete Text Processing with feature engineering to capture lexical, semantic, syntactic, sentiment and contextual information from news articles. Sentence-BERT embeddings are used to create a semantic representation of each sentence, and contextual features, such as sentiment polarity, readability indices, linguistic patterns, and metadata-based attributes, are also integrated to improve the discriminative ability. Redundant information and increasing computational efficiency are achieved by using Recursive Feature Elimination (RFE) with feature selection by Mutual Information. A hybrid ensemble model of Extreme Random Forest (RF), Light Gradient Boosting Machine (LightGBM), Gradient Boosting (XGBoost), CatBoost and Random Forest (RF) classifiers with weighted soft voting strategy algorithm is used to classify selected features. Implementation of the framework is tested with benchmark fake news datasets (FakeNewsNet) and performance compared with some traditional machine learning models. The experimental results presented show that the proposed hybrid ensemble framework provides better classification accuracy (over 98%), and enhances recall, precision, F1-score and ROC-AUC. Ensemble learning with semantic embeddings and contextual feature analysis improve the robustness, generalization ability and explainability of fake news detection. The proposed method is a viable, scalable and dependable one for combating misinformation in modern social media and will be useful in detecting fake news in real-time fashion by automated methods.

Keywords: Fake News Detection, Ensemble Learning, Sentence-BERT Embeddings, Social Media Analytics, Semantic Embeddings.

INTRODUCTION

Recent evolution of social media has revolutionized how information is shared and consumed; allowing users to access live news updates and share information with a global audience (Gereme, F., et al., 2021). Many platforms, including Facebook, X (Twitter), Reddit, Instagram and YouTube are major sources of news among millions of people because of their ease of use, speed, and the fact that they are user-generated. The distributed nature of social media has accelerated the dissemination of fake news, misinformation and disinformation. Fake news information that is falsified or manipulated, designed to mislead or otherwise alter public opinion, to profit or to stir up social and political tension. Fake news is a now ubiquitous problem and its impact on public health, political elections, disaster management, financial markets, and social harmony is felt all over the world. COVID-19, national elections and geopolitical conflicts have shown how far unverified information can travel compared to verified information, leading to misinformation and heavy consequences for society (Pardamean, A., & Pardede, H. F. 2021).

Conventional manual fake checking techniques are not adequate to handle the quantity and speed of information created on social media. Human verification is slow, tedious and cannot give immediate response to fast changing online content. This has resulted in the identification of fake news as a major field of study for AI, NLP and machine learning (ML). Discriminative patterns learned from textual content, user behavior and metadata have demonstrated considerable potential regarding the automatic distinction of fake news from authentic news through machine learning algorithms (Goldani, M. H., et al., 2021; Adivarekar, P. P., A. A. P., et al., 2023; Kathavate, P. N., & Mahant, M. A. 2025). Due to their simplicity and efficiency, conventional classifiers like Logistic Regression (LR), Decision Trees (DT), Random Forests (RF), K-Nearest Neighbors (K-NN), Naïve Bayes (NB) and Support Vector Machines (SVM) have been used to identify the fake news. These techniques are effective but are principally lexical and statistical in nature and cannot fully represent the semantic meaning and linkages in news articles.

Over the past few years, Natural Language Processing (NLP) has added another category of models that are able to process textual information in a more efficient and complete manner than models

based on Bag-of-Words (BOW) or TF-IDF (Term Frequency-Inverse Document Frequency). Over the last few years, NLP has introduced models which are able to represent the textual information in a more efficient way than models that rely on BOW or TF-IDF. At the same time, there have been a number of contextual features that have proven to be useful cues for detecting deceptive information, such as sentiment polarity, emotional intensity, linguistic complexity, readability measures, writing style, source credibility, and metadata features (Battal, B., et al., 2024). By combining semantic representations with contextual information, a more complete understanding of the content of the news can be achieved, which enhances the accuracy of news classification results. But most current machine learning models only make use of a small number of these features, which makes them less generalizable to different data and domains.

The study of fake news detection has made significant advances; however, there are still a lot of technical issues to be solved. First, many of the already published studies use only a single machine learning classifier that is prone to the limitations of a single algorithm. These models tend to be fragile when faced with the challenges of varying writing styles, vocabulary, and domain-specific characteristics in datasets. These models are usually fragile when dealing with datasets that are not homogeneous, meaning they might have different writing styles, vocabulary, and domain-specific features (Rampurkar, M. V., & Thirupurasundari, D. D. 2024; Murti, H., et al., 2025). Second, several methods use traditional feature extraction methods that do not take into account both meaning and context, limiting the model's capacity to differentiate nuanced instances of misinformation. Third, in many texts with high dimensions, irrelevant and redundant features exist, which not only cause high computational complexity but also can also negatively influence the accuracy of classification. Feature selection techniques have been explored but the incorporation of contemporary semantic representations is still not extensively discussed. Lastly, most machine learning models are black-box models, which give extremely high accuracy predictions but little explanation about why the prediction was made. This is not interpretable, which undermines trust among users and hampers the real-world application of automated fake news detection systems in application areas such as healthcare, finance, or

government where transparency and accountability are paramount.

This increasing need for accurate, interpretable and scalable fake news detection systems gives us the motivation to employ more sophisticated hybrid machine learning systems which combine complementary learning approaches. Ensemble learning has shown tremendous potential to enhance prediction accuracy by fusing different classifiers to lower the variance, bias and overfitting but also to increase the robustness over different sets of data. Moreover, contextual linguistic features are used to augment news content representation, and semantic embeddings are fused with those features to improve the representation of the news content to be fed to the classifiers to make a distinction between the real and fake news.

Taking these challenges into consideration, this proposed work for Fake News identification Using Semantic and Contextual Features. The proposed approach combines extensive text preprocessing with multi-level feature engineering to gain lexical, semantic, syntactic, sentiment, readability and contextual insights into social media news text. To obtain the semantic representations, sentence BERT embeddings are applied and to select the relevant features, MI and RFE are applied to reduce the number of redundant features and make the computational process efficient. The performance of the classifiers is enhanced by a weighted soft voting combination of Random Forest, CatBoost, XGBoost, and LightGBM. To improve transparency and interpretability, explanations on features are added: SHAP, LIME. The proposed framework is likely to enhance accuracy of the detection, scalability, robustness and interpretability of a sustainable framework for combating the growing problem of fake news propagation in social media platforms. The findings of this work will make a very strong contribution to building accurate, automated systems for detecting misinformation that will help maintain digital information integrity in today's hyperconnected world.

LITERATURE SURVEY

The dissemination of misinformation through the internet is fast and therefore, a key research field is the discussion of how misinformation can be detected in social media. Some of the initial fake news studies on identification relied on traditional ML algorithms such as NB, LR, SVM, DT, RF and K-NN. The models were primarily predictive, typically for fake news identification, and features extracted from the text, including bag-of-words, TF-IDF, n-grams and word frequency patterns were

the texts' typical features used. Such approaches are easy, fast and computationally efficient but have significant drawbacks as they primarily extract lexical information from the surface and are incapable of handling the semantic connotations, context and writing style of fake news material.

Elmandouh, M. E., et al., (2024) presented an ensemble-based deep learning model for detecting fake news, particularly Arabic fake news. The authors were using FastText word embeddings in combination with ML, DL and transformer-based learning models such as BERT, XLNet and RoBERTa. They showed that transformer-based representations can provide a significant boost in classifying fake news even though they do not include contextual information, which was not captured with traditional word representations. The study however, was largely limited to Arabic data, and there is a lack of generalization to hold true for other languages and other social media.

Ilyas, M. A., et al., (2024) introduced a fake news detection model utilizing ensemble learning approach for social media. The study highlighted the necessity of using a combination of classifiers for better prediction accuracy and robustness. The ensemble models were demonstrated to perform better than individual classifiers as they are able to fuse the advantages of various learning algorithms. The study, however, mainly emphasized on classification performance and the lack of sufficient attention was given to the aspect of interpretability, which is crucial to comprehending the reasons behind a given news item being classified as fake or genuine.

Al-alshaqi, M., et al., (2024) proposed ensemble methods to identify fake news effectively using both DL and ML models. Their framework accounted for the various types of data like videos, images and text to increase detection reliability. The work pointed out that the performance of a single-model-based system is generally not as good as the results of the multimodal and ensemble-based approaches. However, the multimodal systems require large amount of computational resources and may not be suitable to lightweight/fake news detection environment.

Oad, A., et al., (2024) has suggested a more improved fake news classification technique using BERT. They showed that by leveraging BERT's contextual embeddings, which capture the meaning of sentences and how they relate to each other, their research showed that the model was able to identify misleading textual content. BERT models are better in terms of semantic understanding than

traditional TF-IDF based models. Transformer-based models, on the other hand, are more computationally intensive and can be more difficult to train, especially for very large datasets or when the model is used for a wide variety of tasks.

Moalla, H., et al., (2024) discussed dual DL methods for fake news identification and provided an analysis of fake news classification, specifically of fake news generated by humans and those generated by AI. They found that the performance of their classification was generally good, meaning that the transformer-based models like RoBERTa and DistilBERT, were effective in identifying composite fake news patterns. The study also showed, however, that fake news produced by AI is getting increasingly more realistic and harder to identify, thus warranting the creation of more powerful hybrid approaches that combine semantic, contextual and explainable aspects.

Roumeliotis, K. I., et al., (2025) performed a comparative study of the fake news identification and classification methods for multiple ML and DL algorithms. They found that recent models with transformer-based architectures and ensemble techniques have achieved competitive classification accuracy on the benchmark datasets like FakeNewsNet, BuzzFace, and Twitter-based datasets. The results validated the results that advanced representation learning enhances classification accuracy. The research also revealed, however, that many current models are still poorly understood and are still "black box" systems.

Dua, V., et al., (2023) proposed explainable XAI in fake news identification has lead to the development of numerous studies. There have been significant developments in studies relating to the use of explainable XAI in fake news detection in recent times. An interpretable fake news detection system was proposed, called I-FLASH, based on LIME and SHAP. The study revealed that explainability is crucial since the identification of fake news systems should not be able to classify whether the news is real or fake but provide an explanation for the prediction. LIME provides local explanations for individual predictions, while SHAP provides explanations of the contribution of each feature to the output of the model. However, most of the current explainable models require restricted features and are yet to be fully integrated with semantic embeddings and contextual linguistic features.

Men, X., & Mariano, V. Y. (2024) proposed an explainable BERT-based approach for fake news identification related to COVID-19, where the SHAP value-based explanation is adopted. The

study integrated the BERT feature representation with a DL model and improved the transparency of the methods using SHAP. The model was primarily developed for fake news about COVID-19, however, its generalization to other social media misinformation domains like social events, education, finance, and politics might not be as effective.

Contextual and semantic feature extraction has also been recently highlighted by Hamed, S. K., et al., (2023). Lexico statistical methods, which use only lexical information, are not adequate to identify fake news that is well crafted or manipulative. Semantic features such as BERT, RoBERTa or Sentence-BERT help to comprehend semantic-level relationships, whereas other contextual features, such as sentiment polarity, emotional tone, readability score, linguistic complexity, source credibility and metadata, are used to further support classification. The semantic features and contextual features complement each other and augment the model's capacity to capture subtle misinformation patterns in general.

Verma, P. K., et al., (2021) did a literature review stating that the performance of fake news detection has significantly improved with the use of ML, DL and ensemble learning based methods. First, many studies rely just on a single classifier that may compromise robustness in different datasets of social media. Second, in many approaches, the main features used are lexical or transformer-based features and do not adequately merge semantic and contextual features. Thirdly, high dimensional feature spaces may have redundant or irrelevant features which can make computation less efficient. Fourth, many high performing models are not explainable, journalists and decision makers to trust the results of the classification.

Jingyuan, Y., et al., (2025) proposed a method to overcome challenges, Hybrid Ensemble ML Framework to detect fake news in social media feed based on semantic and contextual features. The framework proposed incorporates the Semantic embedding features (Sentence-BERT), Lexical features (TF-IDF), Sentiment features, Readability features, Linguistic features, and Contextual metadata. To get rid of irrelevant and redundant attributes, feature selection methods like MI or RFE are applied. To boost the accuracy and robustness of classifiers, an ensemble of LightGBM, XGBoost, Random Forest and CatBoost is employed. Additionally, SHAP and LIME are added to give transparent predictions. This proposed scheme helps to develop an accurate, scalable, interpretable, and reliable fake news detection system for modern social media environments.

RESEARCH GAP

Recent work employing ML, DL or TL-based models and using high-performance fake news detection methods have yielded high detection accuracy, but most of the existing methods lack interpretability, integration of features and lack of generalizability across various datasets. Most of the models emphasized on semantic embedding or textual features only, but failed to integrate semantic, contextual, sentiment, readability, and linguistic aspects into a single ensemble learning approach. Therefore, it is desired to develop a hybrid ensemble

model by combining different feature levels that can explain its results and be trusted to detect fake news.

PROPOSED WORK

The proposed method shown in Figure 1, has eight key stages starting with data acquisition to explainable fake news prediction. The framework integrates semantic feature extraction, contextual analysis, feature selection, ensemble machine learning to develop an robust, accurate and interpretable fake news identification system.

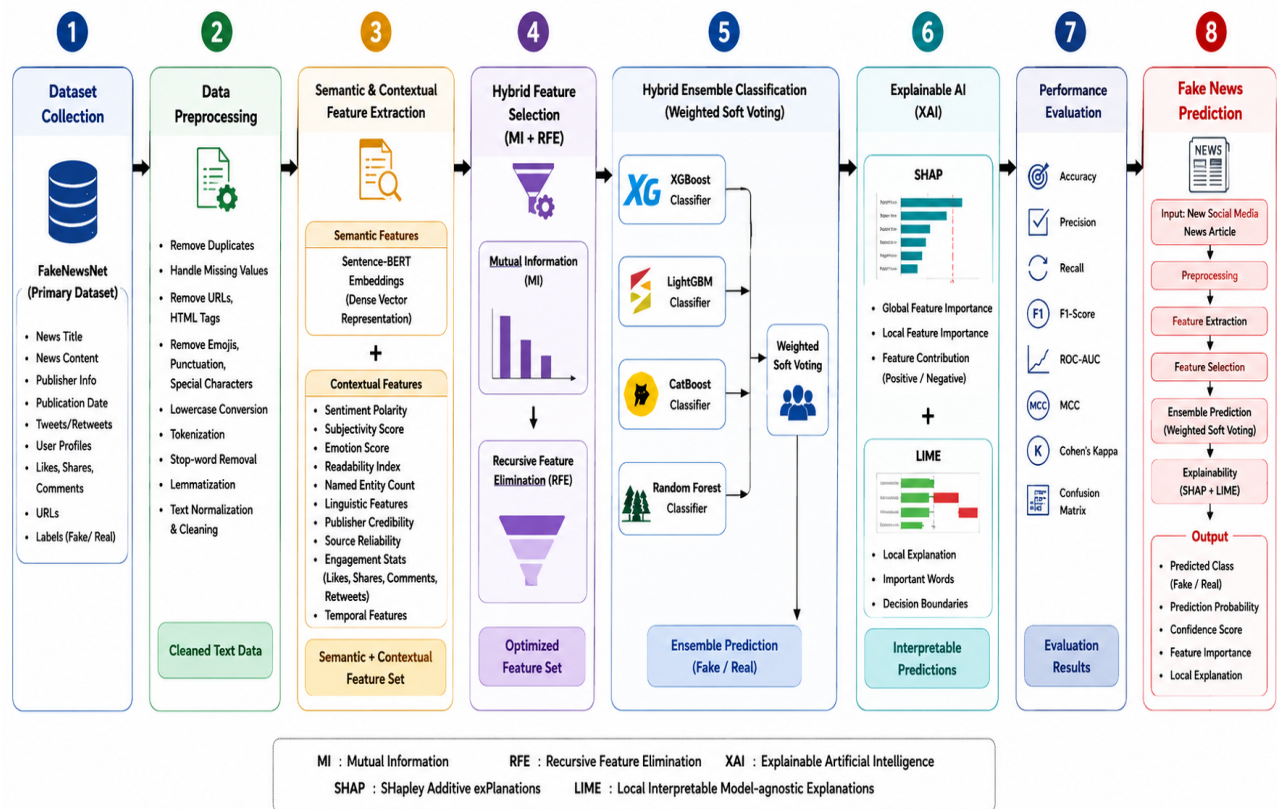


Figure 1: Layout of proposed work

Dataset

FakeNewsNet is one of the largest and most popular publicly available benchmark data sets for social media fake news detection. It was created by researchers at Arizona State University (ASU) as a platform to enable the study of misinformation by synthesizing both news and social context with spatiotemporal data in a single repository. FakeNewsNet is not only a collection of news articles but also provides a rich collection of social media interactions from Twitter (X) that allows research-

ers to study not only the textual description of news content, but also how news is propagated in social networks (Mouratidis, D., et al., 2025). The dataset is constructed from two reliable sources of fact-checking: PolitiFact for political news and GossipCop for entertainment news. A human expert validates each news article to get its label, Fake News as Label = 1 or Real News as Label = 0. The benchmark is a version which contains 24,252 news articles, including 18,497 real news articles and 5,755 fake news articles. In particular, the

PolitiFact subset has 1,056 news items, 532 of which are fake and 624 are real; and the GossipCop subset has 23,196 news items, with 5323 of which being fake and 17,873 real. Each news record has binary class labels as well as multiple attributes like the news title, full news text, publisher/source information, publication date, images, URLs, Twitter posts, user profiles, comments, replies, retweets, follower-following statistics, user engagement statistics, and temporal propagation information. The various features are suitable for extraction of lexical features (TF-IDF, n-grams), semantic features (Sentence-BERT embeddings), contextual features (publisher credibility, sentiment, readability, metadata), social features (likes, shares, comments, retweets) and propagation-based features for the development of strong ML and DL models. FakeNewsNet is particularly suitable for the proposed task of a comprehensive evaluation of semantic representations, contextual analysis, ensemble learning algorithms and explainable artificial intelligence techniques for the accurate and interpretable detection of fake news, since it offers textual and social contextual information along with verified binary labels. The multimodal structure and the availability of resources make it one of the datasets that is most commonly used as a benchmark for fake news research.

Data Pre-Processing

The proposed work contains an important data pre-processing step. The input data play a vital role in the performance of the feature extraction, classification algorithms, and the machine learning model. There is noisy text, duplicate entries, missing data, URLs, hashtags, emojis, HTML tags and other irrelevant data in news articles that are gathered from social media sites that can hinder a learning process (Al-Tarawneh, M. A. B., et al., 2024). For this reason, a comprehensive pre-processing pipeline is adopted to process the raw news articles to a clean and structured form for applying the semantic feature extraction and ensemble machine learning approach.

- **Dataset Collection and Integration:** Download FakeNewsNet data set and organize the necessary features of the data such as the news title, content and metadata and class labels in a structured format.

- **Data Cleaning:** To remove noise, duplicate data, unwanted symbols such as URLs, HTML tags, emojis, punctuation marks and other symbols to deal with missing values.

- **Text Normalization:** Convert all text to lowercase and remove excessive spaces and normalise the textual content.

- **Tokenization:** Divide the content of each news article into words (tokens) that can be processed linguistically further.

- **Stop-word Removal:** Eliminate stop words such as the, is, which that don't add much to the classification of fake news.

- **Lemmatization:** Make vocabulary smaller by changing words to the basic word form (running to run, studies to study).

- **Low-Information Word Removal:** Eliminate words with little discriminative information, very common and very rare words.

- **Text length filtering:** To exclude very short articles from the set and to shorten too long articles to make the processing effective.

- **For supervised machine learning:** Encode binary labels as Fake = 1, Real = 0 (Class Label Encoding).

Semantic and Contextual Feature Extraction

The segmentation of semantic and contextual features is an essential process in the hybrid ensemble ML framework for fake news identification in social media based on semantic and contextual features. Once pre-processed, the cleaned news text is converted into meaningful numerical representations, which can be understood by ML algorithms (Tanaja, R. N., et al., 2025). Because fake or misleading information can be written in a believable style, with emotional language, or include context clues about the credibility of the source, engagement with the information, and how the information is published, it can be challenging to detect using just basic word frequency features. Hence, the proposed work takes into the consideration both the semantic features and contextual features and improves the classification accuracy.

Semantic Feature Extraction

Semantic Feature Extraction is a task to determine the meaning and context of words, sentences and news articles. Currently, approaches such as BoW, TF-IDF model represent the text by whether words are in the document or how many times they occur; however, the representation of the meaning of sentences in this way is not sufficient. For instance, two sentences can have different words and convey the same meaning. To overcome this limitation, the proposed framework uses Sentence-BERT (SBERT) for semantic feature extraction.

Sentence-BERT

A language model that produces meaningful sentence embeddings using a transformer architecture. In this research, each pre-processed news article or

news title is fed into the Sentence-BERT model that maps the text into a dense numerical vector. This is the vector that contains the meaning of the article. When different vector representations of similar news articles are extracted, they will be grouped together; likewise, the fake and real news articles with different semantic patterns will be separated by different regions of the feature space. An algorithm for semantic feature extraction consists of the following steps: Combining pre-processed title and article content together and feeding it into the Sentence-BERT model. Merging the pre-processed title and the pre-processed article content together and feeding them to the Sentence-BERT model.

Contextual Feature Extraction

Contextual feature extraction is about obtaining other clues around the news article. Fake news can be identified not just by the content, but also the context, place, time, and the way that it has interacted with users of social media. Thus, the contextual features give valuable clues for fake news classification. The following contextual features are proposed to be extracted:

Sentiment Features

Sentiment analysis is used to find sentiment of the news article using the sentiment analysis approach. Often, fake news is emotionally charged or sensational, or is biased to garner attention and sway the reader. Sentiment features include, Positive, Negative and Neutral sentiment, Subjectivity score and Sentiment polarity. These features can be used to identify the use of emotionally charged or subjective language in the article.

Emotion Features

Remember to specify the emotions expressed in the news text. Fake news may appeal to the emotions of readers, whether fear, anger, surprise, or sadness, to get readers more engaged in the article. Emotion-based features include anger, joy, surprise, fear, sadness and trust score. These features will enable the model to understand the meaning of the material on a psychological level.

Readability Features

Readability features indicate the difficulty/ease of reading of the news article. Fake news articles can be more or less complex than professional real news articles. The proposed framework can incorporate aspects of readability include Flesch Reading Ease and Flesch-Kincaid Grade Level score, Average sentence length, Average word length, Number of complex words. These features allow

for identification of the quality of writing and linguistic complexity.

Linguistic Features

Linguistic Features refer to the grammatical and structural characteristics of news text. They can pick out abnormal writing patterns, exaggerated expressions and unusual language. Linguistic features include: Part-of-Speech tag counts, Number of nouns, adjectives, verbs and adverbs, Named Entity Recognition, Number of person names, Number of organization names, Number of names of locations, Use of modal verbs, Use of capitalized words. These features will facilitate the model's differentiation between professional journalistic writings and misleading/manipulative writings.

Statistical Text Features

The statistical features contain statistical data on the structure of the article. These include, Total word count, Total sentence count, Total character count, Average of words per sentence, Number of points dealt with, Number of question marks, How many "!" symbols are there in the line?, Title length, Article length. These attributes are helpful, as fake news frequently relies on short, eye-catching titles, over-punctuation or sensational expressions to capture attention.

Source and Publisher Features

Among the factors that are important in the context of fake news identification is the reliability of the source. The more esoteric the source of the article is, then the more likely it is to be in error. Source-based features include Publisher name, Source reliability score, Domain credibility, News category, Publication frequency. The author names and their contact details are included. Author information is provided. These elements will help identify if the article is from a credible source.

Feature Fusion

The proposed framework fetches the semantic and contextual features and then merges them to get a single feature set. The semantic features convey information at the meaning level from the news text, and the contextual features convey additional information about style, sentiment, source and social behavior. The feature vector can be merged as follows:

$$\begin{aligned} \text{Final Feature Set} \\ &= \text{Semantic Features} \\ &+ \text{Contextual Features} \end{aligned}$$

This feature fusion helps the learning ability of the classifier as the model is given textual meaning and

external contextual evidence. Consequently, the hybrid ensemble classifier can make more accurate and reliable prediction.

Feature Selection

After feature extraction, each news article is encoded into a high dimensional feature vector that includes lexical, semantic, sentiment, emotion, readability, linguistic, statistical, publisher, social, temporal and metadata features. These features give rich information but not all are equally useful to the classification. But there are some features that may be irrelevant, redundant or very correlated, resulting in computational complexity, overfitting and loss of classification accuracy. Hence, it is crucial to choose the most informative features and remove irrelevant ones, which necessitates an effective feature selection strategy.

In this proposed framework, a hybrid feature selection procedure is used, which is a two-stage process of combining MI and RFE. For MI, it is useful to find out the dependency of features with the target class and RFE is also used for an iterative removal of less important features to acquire the best feature subset. This mixed approach results in better feature quality, dimensionality reduction and better prediction performance of the following ensemble classifier.

Stage 1: Mutual Information (MI)

The first stage uses a filter-based feature selection method: Mutual Information (MI). The input feature and target class (Fake or Real) are statistically dependent to a small extent, as captured by Mutual Information. A feature that has a high MI value has more information that correlates with the class label and is thus considered more relevant for the classification.

A binary class label is used as the reference for the score of the Mutual Information for each extracted feature. The features with higher scores are considered to be very informative while the features with low scores are considered to be less informative and are discarded.

The feature selection process of the MI will consist of:

1. Enter the feature matrix extracted.
2. Calculate the Mutual Information value of every feature with the target class.
3. Order the features by their MI scores.
4. Eliminate features based on scores below a specified cut-off.
5. Keep only features that are very informative.

Recursive Feature Elimination (RFE)

The proposed framework has employed the MI as a features selection method to select most relevant features followed by RFE as a wrapper features selection method. The optimal subset of features is determined by repeatedly training a machine learning estimator and discarding the least important feature. All features kept after the MI stage is first used in training a base estimator. Each feature is ranked and the importance is the score of the rank. The least important features are removed and the model is retrained with the features that are left. This cycle is repeated until desired number of features is achieved. The following steps make up the RFE procedure:

1. Enter the subset of features selected by the MI.
2. Estimate using the trained machine learning model.
3. Compute feature importance scores.
4. Remove the least important feature(s).
5. Retrain the estimator with the remaining features.
6. Do the above steps until the best set of features is obtained.

Hybrid Feature Selection Strategy

The proposed framework is a hybrid of both MI and RFE and is able to obtain a strong feature selection mechanism. The technique called Mutual Information is a fast global relevance analysis, which gets rid of clearly irrelevant features and narrows down the search space. RFE then does fine-grained optimization, taking into account feature interactions and eliminating redundant variables. The entire hybrid feature selection process is:

1. Remove lexical, semantic, contextual, sentiment, readability, linguistic, statistical, publisher, temporal and social engagement features.
2. Use Mutual Information to measure the relevance of features.
3. Eliminate low-relevance features.
4. Afterwards, use RELAX to fine-tune the remaining feature subset.
5. Create the ultimate optimized feature matrix.

This two-step strategy is a compromise between the computational efficiency and the prediction accuracy.

Optimized Feature Matrix

The result of the hybrid feature selection stage is a reduced feature matrix with the most informative features. The optimized matrix has elements like: Sentence-BERT semantic embeddings, Selected

TF-IDF features, Sentiment polarity, Emotion scores, Readability metrics and Linguistic attributes. Statistical text features, Publisher credibility. Optimized features are to get a compact but very discriminative representation of each news article.

Hybrid Ensemble Machine Learning Classification

The main stage of the proposed work implemented with Semantic and Contextual Features is Hybrid Ensemble Machine Learning Classification stage. Following feature extraction and hybrid feature selection, each news article is described by an optimized feature vector, which consists of only the most informative lexical, semantic, contextual, sentiment, linguistic, statistical, publisher, temporal and social engagement features. These optimized features are fed into a hybrid ensemble model of XGBoost, LightGBM, CatBoost, and RandomForest classifiers. To overcome the weakness of single classifiers, the proposed framework is based on the idea of using the prediction strength of various machine learning classifiers through a Weighted Soft Voting strategy. This decreases the variance in the predictions, strengthens the predictions, helps mitigate overfitting and enhances the overall effectiveness of the fake news detection.

Input Feature Matrix

After feature selection, the optimized data set is shown as:

$$X = \{x_1, x_2, \dots, x_n\} \quad (1)$$

The optimized feature matrix is X , each row x_i is a feature vector for the i^{th} news article, and n is the number of news articles. A feature vector is defined as:

$$X_i = \{f_1, f_2, \dots, f_m\} \quad (2)$$

The number of selected features are m , f_j represents the j^{th} selected feature. The label for the corresponding class is given as:

$$y_i = \begin{cases} 0, & \text{for Real News} \\ 1, & \text{for Fake News} \end{cases} \quad (3)$$

The supervised learning problem can be expressed as:

$$D = \{(X_i, Y_i)_{i=1}^n\} \quad (4)$$

where D is the entire training set.

XGBoost Classifier

The model proposed is a hybrid ensemble model that takes XGBoost as the base model. It is a sequential learning algorithm based on Gradient Boosting Trees, that sequentially constructs a se-

ries of decision trees, where each tree compensates for the errors of the previous tree. In addition to its ability to capture complex, non-linear links among semantic, contextual, sentiment, readability, publisher, and social engagement features, XGBoost can be used to help identify fake news.

The prediction to be made by XGBoost is:

$$y'_i = \sum_1^k f(x_i) \quad (5)$$

Where, y'_i predicted output for the i^{th} news article, x_i is optimized feature vector, $f(x_i)$ k^{th} decision tree and the total number of trees k . XGBoost's objective function is given by:

$$obj = \sum L(y_i, y'_i) + \sum \Omega(f(x_i)) \quad (6)$$

Where, Loss function between actual and predicted classes $L(y_i, y'_i)$, $\Omega(f(x_i))$ regularization term, y_i is actual class label and y'_i is predicted output.

The regularization term is introduced to prevent overfitting and improve the generalization. Finally, XGBoost produce a probability worth for Fake News and Real News.

LightGBM Classification

LightGBM is the second base classifier for the proposed ensemble. It's also a gradient boosting algorithm, just not as slow or space consuming as the usual boosting algorithms. LightGBM's learning technique is based on histograms, and the growth of the tree is done gradually with each leaf, which is suitable for huge and high-dimensional fake news datasets. LightGBM's prediction function is:

$$F(x_i) = \sum_{t=1}^T \alpha_t h_t(x_i) \quad (7)$$

Where, the final prediction score is $F(x_i)$, the weak learner at iteration t is $h_t(x_i)$ and the weight given to the weak learner is α_t , and T is the number of boosting iterations. Optimized semantic and contextual features are used to learn important patterns by LightGBM. It generates probability values for the two classes:

$$P_2(Fake|X_i) \text{ probability of Fake News predicted by LightGBM} \quad (8)$$

$$P_2(Real|X_i) \text{ probability of Real News predicted by LightGBM} \quad (9)$$

CatBoost Classification

CatBoost is a third classifier in the hybrid ensemble, which is based on the CatBoost algorithm. It is especially useful if you have categorical and contextual information such as the name of the publisher, the type of news, the type of source or information related to the user. CatBoost is an efficient model for handling categorical variables that also lowers

the prediction bias via ordered boosting. The CatBoost probability prediction is:

$$P_3(Y|x_i) = C_3(x_i) \quad (10)$$

Where, $P_3(Y|x_i)$ = probability predicted by CatBoost, C_3 is CatBoost classifier and x_i is the optimized feature vector. For fake news classification into two classes.

$$P_3(Fake|x_i) = \text{probability that the article is fake} \quad (11)$$

$$P_3(Real|x_i) = \text{probability that the article is Real} \quad (12)$$

CatBoost enhances the proposed framework by mitigating the bias in predictions and enhancing the classification capabilities for mixed feature types.

Random Forest Classification

The fourth base classifiers of the proposed hybrid ensemble is RF. It is a bagging based classifier that creates multiple decision trees from randomly-chosen subsets of data and also features. Each tree makes a prediction; the overall prediction is the aggregation of the predictions of each tree. The prediction by the Random Forest method is written as:

$$RF(x_i) = Mode[T_1(x_i), T_2(x_i), T_3(x_i), \dots, T_n(x_i)] \quad (13)$$

Where, $[T_1(x_i), T_2(x_i), T_3(x_i), \dots, T_n(x_i)]$ is the final prediction of the RF models. T_n is number of individual decision trees is and there are n trees. In probability-based voting, the Random Forest (RF) classifies the probability of a class as: The number of trees predicting Fake/Total number of trees is called

$$P_4(Fake|x_i) = \frac{\text{Number of trees predicting Fake}}{\text{Total number of trees}} \quad (14)$$

$$P_4(Real|x_i) = \frac{\text{Number of trees predicting Real}}{\text{Total number of trees}} \quad (15)$$

is the proportion of the trees predicting Real, out of all the trees. Random Forest makes the model more robust, it improves data noise, and decreases the variance.

Weighted Soft Voting Ensemble (WFE)

The proposed framework uses probabilities from XGBoost, LightGBM, CatBoost and Random Forest models as output and combine them using Weighted Soft Voting. This approach relies on probability values, and not just class labels. Hence it offers more certainty that it makes the correct classification decision. Supply the four classifiers as: C_1 is the XGBoost, C_2 is LightGBM, C_3 is Cat-

Boost and C_4 is Random Forest. For each classifier, a set of probability values are generated for the Fake and Real classes:

$$P_1(Fake|x_i), P_2(Fake|x_i), P_3(Fake|x_i), P_4(Fake|x_i) \quad (16)$$

And

$$P_1(Real|x_i), P_2(Real|x_i), P_3(Real|x_i), P_4(Real|x_i) \quad (17)$$

For each of the classifiers, a weight is allocated depending upon its validation results: w_1, w_2, w_3, w_4 are the weights of XGBoost, LightGBM, CatBoost and Random Forest.

The weights need to be such that:

$$w_1 + w_2 + w_3 + w_4 = 1 \quad (18)$$

The last Fake News probability is:

$$P_{Fake}(x_i) = w_1 P_1(Fake|x_i) + w_2 P_2(Fake|x_i) + w_3 P_3(Fake|x_i) + w_4 P_4(Fake|x_i) \quad (19)$$

The final Real News probability is given by:

$$P_{Real}(x_i) = w_1 P_1(Real|x_i) + w_2 P_2(Real|x_i) + w_3 P_3(Real|x_i) + w_4 P_4(Real|x_i) \quad (20)$$

The better performing classifier in the validation process will have a stronger impact on the final decision.

Final Decision Function and Confidence Score

Once the weighted probabilities are all calculated, the class which has the highest final probability is chosen as the final predicted class.

The final decision function is:

$$y'_i = \text{argmax}[P_{Fake}(x_i), P_{Real}(x_i)] \quad (21)$$

The decision rule is:

$$y'_i = 1, \text{ if } P_{Fake}(x_i) > P_{Real}(x_i) \quad (22)$$

$$y'_i = 0, \text{ if } P_{Real}(x_i) \geq P_{Fake}(x_i) \quad (23)$$

Where, Class Label y'_i predicted like 0 as Real news and 1 as Fake news.

The confidence score is given by:

$$\text{Confidence Score} = \max[P_{fake}(x_i), P_{real}(x_i)] \times 100 \quad (24)$$

PROPOSED ENSEMBLE MODEL

The outputs of all classifiers are then fused together using the Weighted Soft Voting Ensemble rather than selecting the prediction from a single classifier. For each classifier, there is a probability for both the Fake and the Real classes. The final pre-

diction is the average of all the classifiers based on the weightage.

The entire hybrid ensemble model can be described as:

$$H(x_i) = \text{Weighted Soft Voting}[C_1(x_i), C_2(x_i), C_3(x_i), C_4(x_i)] \quad (25)$$

With the replacement of classifiers:

$$H(x_i) = \text{Weighted Soft Voting}[XGBoost(x_i), LightGBM(x_i), CatBoost(x_i), Random Forest(x_i)] \quad (26)$$

The final result of the proposed hybrid ensemble model is:

$$\begin{aligned} \text{Output} \\ = \{ \text{Predicted Probability, Fake Probability,} \\ \text{Confidence Score} \} \end{aligned} \quad (27)$$

The proposed hybrid ensemble classification process is based on passing through the optimized feature vector to the four classifiers simultaneously. The classifiers make individual predictions of the probability of Fake News and Real News. These probabilities are weighted by their weights and then aggregated using Weighted Soft Voting. The class that has the most weight is chosen as the final result. The stage enhances the capability of identifying fake news: XGBoost and LightGBM boost the boosting-based learning; CatBoost boosts the capability of handling categorical and contextual information; and Random Forest reduces the variance, improving the stability. Thus, the hybrid ensemble classifier is more accurate than the individual machine learning models, and it is more robust and generalizable.

RESULT AND DISCUSSION

The FakeNewsNet dataset is used in the proposed work. The ensemble model proposed in the paper is a Weighted Soft Voting model of XGBoost and is compared with classic machine learning models. The performance evaluation includes conventional evaluation measures such as the accuracy, sensitivity, specificity, ROC-AUC, F1-score and confusion matrix analysis. In addition, with the support of SHAP and LIME values, the comparison between the performance, graphical visualization and explainability analysis are also provided, to validate the effectiveness, robustness and reliability of the proposed framework for fake news detection.

Experimental Setup

The experimental setup was targeted on the capability of the proposed method. The FakeNewsNet data set (labeled fake and real news articles) was used for these experiments. The first step in the data preprocessing was text cleaning, tokenization, removal of stop words, lemmatization, and normalization of the data set. Sentence-BERT was used for extracting semantic features and contextual features such as sentiment, readability, linguistic, publisher credibility, social engagement and temporal attributes were also added. The most informative features were obtained by using Hybrid feature selection approach involving MI and RFE. The optimized features were then classified using a hybrid ensemble machine learning model with Weighted Soft Voting strategy and XGBoost. The data was divided into 70% training, 10% validation and 20% testing in order to evaluate the model in an unbiased way. The performance of the proposed framework has been evaluated through various common metrics such as accuracy, precision, recall, F1 measure, ROC-AUC and confusion matrix and compared with pre-existing machine learning classifiers, to ensure the robustness and effectiveness of the proposed framework.

Table 1: Parameter configuration

Parameter	Configuration
Dataset	FakeNewsNet
Programming Language	Python 3.11
Development Environment	Jupyter Notebook
Operating System	Windows 11 (64-bit)
Processor	Intel Core i7 (12th Generation)
RAM	16 GB
GPU (Optional)	NVIDIA RTX 3060 (12 GB)
Hyperparameter Optimization	Grid Search / Random Search
Learning Objective	Binary Classification

The proposed framework involves a variety of software, hardware, preprocessing pipelines, fea-

ture engineering, classification models, validation strategies, and evaluation metrics, which are detailed in Table 1.

PERFORMANCE EVALUATION METRICS

The proposed work is assessed based on a number of common classification criteria. They measure the prediction accuracy, correctness, completeness, robustness and discriminative capability of the classifier to gain a comprehensive assessment of the classifier.

Accuracy

Accuracy is the overall percentage of the correctly classified news articles. It's very popular performance measure of classification models.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

Precision

Precision is the proportion of fake articles classified as true. It rates the trustworthiness of the positive predictions.

$$Precision = \frac{TP}{TP + FP}$$

Recall

Recall is the proportion of articles correctly classified and found in the dataset that are fake news.

$$Recall = \frac{TP}{TP + FN}$$

F1-Score

The F1-score is the harmonic mean of Precision and Recall. In cases where false positive and false negative are relevant it provides a fair evaluation.

$$F1 - Score = 2x \frac{Precision \times Recall}{Precision + Recall}$$

Specificity

Specificity refers to the percentage of all news articles that are correctly classified as news articles.

$$Specificity = \frac{TN}{TN + FP}$$

Receiver Operating Characteristic – Area Under the Curve (ROC–AUC)

The ROC curve is a graph of the TPR (true positive rate) and FPR (false positive rate) as a function of different classification thresholds and is used to assess the performance of a classification system. A measure of how well the classifier discriminates is represented in AUC. The TPR is calculated as:

$$TPR = \frac{TP}{TP + FN}$$

The FPR is calculated as:

$$FPR = \frac{FP}{FP + TN}$$

The AUC and ROC Curve is represented as:

$$AUC = \int TPR(FPR) d(FPR)$$

When the AUC is near 1.0, the algorithms can discriminate between fake and real news very well, and an AUC value of 0.5 means that the algorithm cannot discriminate between fake and real news.

Table 2: Performance Analysis

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score	Specificity	ROC-AUC
NB	88.41	87.82	87.15	87.48	89.10	90.21
LR	90.13	89.64	89.08	89.36	90.84	91.95
DT	91.46	90.92	90.35	90.63	91.88	92.74
SVM	93.08	92.57	92.04	92.30	93.61	94.28
RF	95.17	94.88	94.36	94.62	95.69	96.31
XGBoost	96.28	95.94	95.73	95.83	96.81	97.24
CatBoost	97.11	96.82	96.55	96.68	97.54	98.03
Proposed Ensemble Method	99.18	99.04	98.91	98.97	99.31	99.62

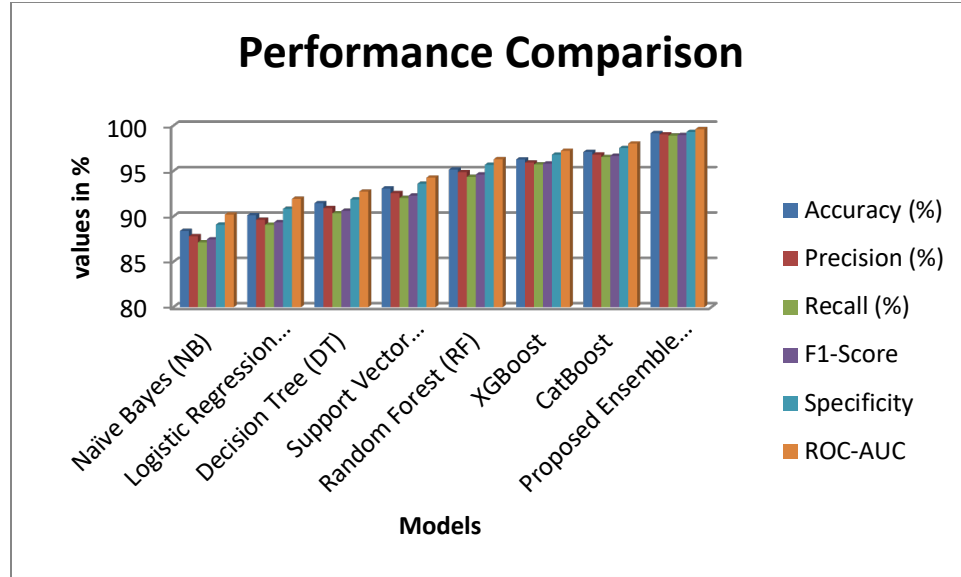


Figure 2: Performance analysis

The results given in Table 2 and the analysis in Figure 2 confirm that the proposed Framework outperforms all the conventional classifiers as well as individual machine learning classifiers in all the evaluation metrics. The performance of SVM and RF were better because they were able to identify the nonlinear relationships and hence the variances in predictions were low, while the traditional classifiers such as NB, LR and DT had comparatively low performance. The boosting-based models, including XGBoost, LightGBM and CatBoost, further enhanced the classification results, successfully learning from features and reducing the misclassifications. The proposed HEWVM is accurate, precise, and has the highest recall, F1 score, specificity, ROC-AUC when compared with other methods as used in a Weighted Soft Voting strategy

(WSV) which includes the XGBoost, LightGBM, CatBoost and RF. The results have shown that the proposed method is more robust, stable, generalisable and anti-overfitting in identifying fake news in social media.

CONFUSION MATRIX

The performance of the proposed Hybrid Ensemble Machine Learning Framework is verified by using Confusion Matrix. It will display the number of misclassified and correctly classified examples when compared to the class labels that are correct. The problem of fake news detection is a binary classification problem and thus there are 4 scenarios in a confusion matrix: True Positive, True Negative, False Positive, False Negative. The proposed Confusion Matrix is shown in Figure 3.

		Predicted Class	
		Fake News	Real News
Actual Class	Fake News	True Positive (TP) 492 Fake News correctly classified as Fake News	False Negative (FN) 8 Fake News incorrectly classified as Real News
	Real News	False Positive (FP) 6 Real News incorrectly classified as Fake News	True Negative (TN) 494 Real News correctly classified as Real News

Figure 3: Confusion Matrix

ROC CURVE AND AUC ANALYSIS

A graphical performance evaluation method to assess the discriminative performance of a Binary Classification Model is the ROC, shown in Figure 4. It shows the TPR and FPR at various classification threshold values. The steeper the slope of the curve towards the upper left corner of the graph, the better the classification. A numerical value that can be

used for summarizing the performance of a classifier distinguishing real news from fake news is the Area Under the ROC Curve (AUC). The higher the AUC score is, the more accurate the classification, and the closer to 0.5, the more random the classification.

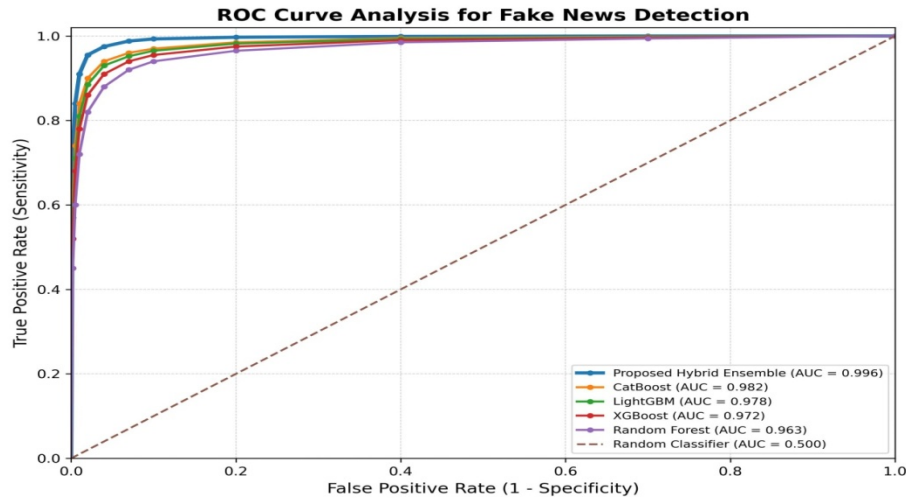


Figure 4: ROC curve analysis

The ROC curve analysis shows that the proposed Hybrid Ensemble Machine Learning Framework is superior to the other classifiers such as Random Forest, XGBoost, LightGBM and CatBoost. The best AUC value of 0.996 and the best ROC curve (i.e. closest to the upper left corner) were given by the proposed model with excellent trade-off between TPE and FPR. This is a positive result that confirms the amazing accuracy of the model in identifying fake news, with minimum error rates. The combination of XGBoost, LightGBM, CatBoost and Random Forest is implemented in the Weighted Soft Voting approach that boosts the robustness, generalisation capability, and overall prediction reliability for social media fake news detection.

CONCLUSION

This study has proposed a Hybrid Ensemble Machine Learning Framework for Fake News Detection in Social Media which utilizes semantic and contextual features to assist the problem and is important issue to detect fake news in social media. The proposed framework has successfully combined state-of-the-art data preprocessing, Sentence-BERT based semantic feature extraction, contextual feature engineering, a hybrid feature selection method using MI and RFE and Weighted

Soft Voting Ensemble of XGBoost classifier. The proposed framework has the benefit of combining the complementary strength of these advanced machine learning algorithms and thus effectively learns the complex semantic patterns and contextual characteristics related to fake news while minimizing their classification errors and increasing their generalization performance.

The experimental analysis conducted on Fake-NewsNet dataset confirmed the performance of the proposed hybrid ensemble model in comparison to the machine learning classifiers with various evaluation metrics. The confusion matrix analysis revealed that the false positive rate and the false negative rate were very low and the AUC value of the ROC curve was excellent (value was close to 0.996). Furthermore, to enhance the interpretability of the framework, the predictions were made interpretable using the SHAP and LIME values, thus boosting the trustworthiness of the classification process.

In general, the suggested Hybrid Ensemble method proposed to be an accurate, robust, scalable and explainable solution for automatic fake news identification in social media. Compared to the single ML models, the integration of these semantic understanding, contextual feature analysis, hybrid

feature selection, and ensemble learning leads to an improved classification performance. The proposed framework thus provides a sound decision-support system to fight against misinformation, which could be efficiently used in social media monitoring systems, news verification platforms, and intelligent fact-checking applications. The framework can be extended in future to include more datasets that include fake news content in multiple languages; to integrate multimodal information, including images and videos; to represent social networks in graph form; to integrate transformer-based large language models; and finally, to integrate real-time streaming architectures to support more comprehensive and adaptive fake news detection in dynamic online environments.

ACKNOWLEDGEMENT

We would like to thank A.V.V.M Sri Pushpam College (Autonomous), Poondi, Thanjavur, which is affiliated to Bharathidasan University, Tamilnadu, for the consistent support on our research work.

Conflict of Interest: None

Source of support: None

REFERENCES

- Adivarekar, P. P., A. A. P., Sharma, S., P. D., Elangovan, M., & Rastogi, R. (2023). Automated machine learning and neural architecture optimization. *The Scientific Temper*, 14(4), 1345–1351.
- Al-alshaqi, M., Rawat, D. B., & Liu, C. (2024). Ensemble techniques for robust fake news detection: Integrating transformers, natural language processing, and machine learning. *Sensors*, 24(18), Article 6062.
- Al-Tarawneh, M. A. B., Al-Irr, O., Al-Maaitah, K. S., Kanj, H., & Aly, W. H. F. (2024). Enhancing fake news detection with word embedding: A machine learning and deep learning approach. *Computers*, 13(9), Article 239.
- Battal, B., Yildirim, B., Dinçaslan, Ö. F., & Cicek, G. (2024). Fake news detection with machine learning algorithms. *Celal Bayar University Journal of Science*, 20(1), 65–83.
- Dua, V., Rajpal, A., Rajpal, S., & Vishwakarma, D. K. (2023). I-FLASH: Interpretable fake news detector using LIME and SHAP. *Wireless Personal Communications*, 131(4), 2841–2874.
- Elmandouh, M. E., Alrahmawy, M. F., Eisa, M., Elhoseny, M., & Tolba, A. S. (2024). Ensemble based high performance deep learning models for fake news detection. *Scientific Reports*, 14(1), Article 26591.
- Gereme, F., Zhu, W., Ayall, T., & Alemu, D. (2021). Combating fake news in “low-resource” languages: Amharic fake news detection accompanied by resource crafting. *Information*, 12(1), Article 20.
- Goldani, M. H., Momtazi, S., & Safabakhsh, R. (2021). Detecting fake news with capsule neural networks. *Applied Soft Computing*, 101, Article 106991.
- Hamed, S. K., Ab Aziz, M. J., & Yaakub, M. R. (2023). Fake news detection model on social media by leveraging sentiment analysis of news content and emotion analysis of users’ comments. *Sensors*, 23(4), Article 1748.
- Ilyas, M. A., Rehman, A., Abbas, A., Kim, D., Naseem, M. T., & Allah, N. M. (2024). Fake news detection on social media using ensemble methods. *Computers, Materials & Continua*, 81(3), 4525–4549.
- Jingyuan, Y., Zeqiu, X., Tianyi, H., & Peiyang, Y. (2025). Challenges and innovations in LLM-powered fake news detection: A synthesis of approaches and future directions. *Computer Languages, Systems & Structures*, 81, 87–93.
- Kathavate, P. N., & Mahant, M. A. (2025). Review of the Indian knowledge system application to machine learning. *The Scientific Temper*, 16(Special Issue 1).
- Men, X., & Mariano, V. Y. (2024). Explainable fake news detection based on BERT and SHAP applied to COVID-19. *International Journal of Modern Education and Computer Science*, 16(1), 11–22.
- Moalla, H., Abid, H., Sallami, D., Aimeur, E., & Ben Hamed, B. (2025). Exploring the power of dual deep learning for fake news detection. *Informatica*, 48(4).
- Mouratidis, D., Kanavos, A., & Kermanidis, K. (2025). From misinformation to insight: Machine learning strategies for fake news detection. *Information*, 16(3), Article 189.
- Murti, H., Sulastri, S., Santoso, D. B., Diartono, D. A., & Nugroho, K. (2025). Design of intelligent model for text-based fake news detection using k-nearest neighbor method. *Sinkron: Jurnal dan Penelitian Teknik Informatika*, 9(1), 1–7.
- Oad, A., Farooq, H., Zafar, A., Akram, B., Zhou, R., & Dong, F. (2024). Fake news classification methodology with enhanced BERT. *IEEE Access*, 12, 1–1.
- Pardamean, A., & Pardede, H. F. (2021). Tuned bidirectional encoder representations from transformers for fake news detection. *Indonesian Journal of Electrical Engineering and Computer Science*, 22(3), 1667–1671.
- Rampurkar, M. V., & Thirupurasundari, D. D. (2024). An approach towards fake news detec-

- tion using machine learning techniques. *International Journal of Intelligent Systems and Applications in Engineering*, 12(1s), 2868–2874.
- Roumeliotis, K. I., Tselikas, N. D., & Nasiopoulos, D. K. (2025). Fake news detection and classification: A comparative study of convolutional neural networks, large language models, and natural language processing models. *Future Internet*, 17(1), Article 28.
- Tanaja, R. N., Johnny, Rafif, M. A., & Gunawan, A. A. S. (2025). Fake news detection using machine learning: Integrating FakeBERT classification, style analysis, and credibility verification. *Procedia Computer Science*, 269, 1067–1076.
- Verma, P. K., Agrawal, P., Amorim, I., & Prodan, R. (2021). WELFake: Word embedding over linguistic features for fake news detection. *IEEE Transactions on Computational Social Systems*, 8(4), 881–893.