



Research Article,

An Intelligent Framework for Automated Fake News Detection

Ayesha Maryam^{1,*}, Faisal Shazad¹

¹ Independent Researcher of Computer Science, Multan, 59100, Pakistan

² Department of Emerging Computing Technologies, Emerson University, Multan, 60000, Pakistan

*Corresponding Author: Ayesha Maryam. Email: ayeshamaryam2133@gmail.com

Received: 10 June 2026; Revised: 17 July 2026; Accepted: 03 August 2026; Published: 15 August 2026;

AID: 005-02-000070

<https://doi.org/10.66108/mna.v5i2.131>

Abstract: As digital news platforms and social media grow and multiply, misinformation has become a widespread problem and automatic methods of detecting fake news are becoming more important. This paper introduces a machine learning-based system to detect fake news by applying traditional classification techniques and Text Representation method Term Frequency-Inverse Document Frequency (TF-IDF). The experiments were carried out on the ISOT Fake and True News dataset which has 44,898 news articles including 23,481 fake news and 21,417 real news articles. The articles were pre-processed for each of the articles to convert the title and body in to lower case and then to exclude all the HTML elements, Punctuation, numbers and the stop words from the articles. TF-IDF (with unigram, bigram) was used for the representation of the resulting text. To train the classifiers, the data was split into an 80:20 ratio to develop the set and train the classifiers. The classifiers were trained by dividing the data set into training and testing sets in the ratio of 80:20. The accuracy, precision, recall, F1 score, and confusion matrix analysis were used to evaluate their performance. Among the tested models, precision, recall and F1 Score were 99.79%, 99.91% and 99.85% respectively with test accuracy of 99.86% with XGBoost. The study results reveal that TF-IDF based text representation along with ensemble learning can be a useful method to classify fake news on the chosen data set. As such, the study proposes an automated fake news detection using a machine learning framework with the adoption of traditional classification methods and ensemble techniques.

Keywords: Fake News Detection; Machine Learning; TF-IDF; Text Classification; Misinformation Detection; XGBoost.;

1. Introduction

The Internet and social media have significantly altered the nature of information production, distribution and consumption. People can easily access news or information via online news portals, social networking sites, blogs, messaging websites, in just a single day. The online news portals, social networking, blogs, messaging websites etc. vouchers to reach huge masses in just a day with news and information delivery. While this fast-sharing of information facilitates information, it has been a facilitator to the increased spread of misinformation and fake news.

Fake news is a term used to describe false information or information that is sensationalized or skewed to drive certain opinions and/or gain financial or social/political ends. The swift dissemination of this type of information can impact a range of democratic processes, public health, the financial economy and emergency response operations. The trust in automated and reliable tools for detecting misleading information spread via digital platforms has been further strengthened by events like the Covid-19 pandemic and political elections.

The widely utilized fact-checking protocols the news media employs can offer reliable verification, but manually making this verification for the vast amounts of information online can be time consuming and thus hard to scale. Recently, there has been a growth of interest in applying machine learning (ML) and natural language processing (NLP) methods to automatically classify fake news and real news articles [1-3]. These methods discover language and statistical structures from labeled data sets and have been proven to be applicable to the task of text classification.

One of the widely used methods to transform the text content into a numerical representation with moderately low computational complexity is the Term Frequency-Inverse Document Frequency (TF-IDF). Furthermore, a number of classifiers for fake news classification, like Random Forest, XGBoost, Logistic Regression, Gradient Boosting, Decision Tree, and Passive Aggressive Classifier, have been used for supervised classification. They can be affected by various aspects of text preprocessing, feature representation, parameter selection, and measures used for the classification. [1-3]

While there have been many studies on fake news detection techniques using automated systems, most of the studied techniques do not consistently compare across a larger slate of commonly used classical machine learning classifiers. Specifically, using a number of different models with the same pre-processing, TFIDF representation, dataset and experiment conditions could give a consistent background for analysis of their classification performance.

To meet the research need, this study proposes a framework for the detection of fake news which involves feature extraction for text using TF-IDF along with six supervised machine learning classifiers. The framework first preprocesses the news content, extracts the unigram and bigram features for the news articles using TF-IDF matrix, and performs experiments and compare the selected classifiers on the basis of common conditions. To explore accuracy, precision, recall, F1 score, and confusion matrix analysis techniques are used to analyze the classification accuracy of the models and determine the model with the best accuracy in the selected dataset.

These are summarized as the main contributions of this study:

- Designing a fake news checker system based on classical machine learning techniques.
- Rather Principle of Gate Particle Processing for News articles before Feature Representation.
- Performance comparison of six supervised machine learning classifiers on the same set of experimental conditions.
- Multiple evaluation measures and confusion matrix analysis for assessing the performance of classifiers.
- Identification of most successful classifier for data set chosen, from experimental results.

The rest of this paper will proceed as follows. Section 2 shows the review of the previous studies related to fake news detection. The dataset and exploratory data analysis is described in Section 3. The pre-processing and feature extraction processes are discussed in Section 4. Section 5 describes the methodology which includes the machine learning models selected. The evaluation metrics are explained in section 6. The experimental results are presented and discussed in Section 7, Section 8 concludes the paper and directions for future research are suggested.

2. Related Work

The problem of detecting fake news has gained momentum as fake news spreads at a fast pace via online news platforms and social media sites. Several studies have introduced various methods for automatically

distinguishing true from fake news articles. Numerous methods have been introduced to automatically classify news articles as true or fake using machine learning, deep learning and transformer-based methods.

The initial investigations used mostly the classical machine learning algorithms and feature extraction methods like Bag-of-Words and TF-IDF. Jouhar et al. proposed a fake news detection framework based on TF-IDF features and various ML classifiers including decision tree, gradient boosting, logistic regression, random forest, passive aggressive and XGBoost. They conducted experiments and found that XGBoost model performed best in terms of highest classification accuracy out of various models studied, which proved the effectiveness of ensemble learning in fake news detection [4].

There are a few researchers who have been comparing the classical machine learning algorithms for fake news detection. Sharma et al. compared the accuracy of Naïve Bayes, Random Forest, and other supervised learning algorithms in detecting misinformation in news articles. They demonstrated that ensemble methods typically perform better than any individual classifiers when they are used under good feature representations. [5]. Likewise, Khanam et al. explored the boosting classifiers such as XGBoost, Support Vector Machine (SVM), and other classifiers, and concluded that the boosting-based classifiers are performing well on the benchmark fake news datasets [6]. It is important to choose appropriate classifiers for detecting textual misinformation and Pandey et al. compared KNN, SVM, Decision Tree and other machine learning methods [7].

The ISOT Fake and True News data set has gained widespread popularity for use as a benchmark data set for fake news detection due to having titles, article text, subjects, and publication dates of fake and real news. Ahmed et al. showed TF-IDF features and n-grams successfully include linguistic characteristics which differentiate fake and factual news articles [8, 9]. Such research founded TF-IDF as one of the ideal methods of baseline feature extraction for the classic text classification problem.

Recent papers have shown that transformer models, such as BERT and its variants, tend to be more effective on a variety of fake news datasets than traditional machine learning approaches. However these models have high computational complexity and training time compared to traditional classifiers. As a result, classical machine learning algorithms have gained wide attention for applications that need to be accessible and efficient [2, 3]. Learning the contextually-based semantic representation directly from the text has yielded a variety of models that have outstanding performance, like CNNs, LSTMs, BERT, RoBERTa, and FakeBERT. These methods have shown better performance than usual machine learning methods, but usually consume significantly more computational power, training time, and data. Thus, classical machine-learning algorithms continue to be interesting in the context of allowing faster training of models, reduced computation requirements, and increased interpretability, for applications where the former is required or the latter is preferred [2, 3, 5, 6, 10-14].

Lately, text-based fake news detection has also been shown through research to be applicable to classical machine learning methods regardless, with the use of effective feature representations like TF-IDF. In-depth review of machine learning methods for fake news detection reveals Logistic Regression, Support Vector Machine, Naïve Bayes, Decision Tree, Random Forest, Gradient Boosting and XGBoost methods as the most commonly explored traditional methods; and the ongoing applications of TF-IDF was also reported for content-based classification [15]. Combination of TFIDF and classic classifiers has also been investigated in recent empirical studies, with comparable classification results in various experimental scenarios and datasets [16, 17]. Other recent studies have compared classical machine learning and deep learning, and transformer-based approaches, where the performance of the models depends on the type of data, features represented, and the experimental setup—without a single modelling technique consistently outperforming the others [18], [19]. The results justify the use of the different classical classifiers in the current work, and also highlight the need to test them using similar experimental conditions.

2.1. Research Gap

While machine learning and deep learning techniques have been extensively studied for fake news detection, some of the reported outcome details deviate in data sets, the pre-processing techniques,

representations of the features, configurations of classifiers and procedures for assessing the results. Thus, the performance of the different models cannot be easily compared directly between studies. More recently, it has been shown that classical and deep learning techniques can have different performances depending on the experimental configuration and the dataset. Thus, a comparatively controlled test for the different classical classifiers to examine their relative performance in the same dataset, preprocessing, TF-IDF representation, train-test split, and measure of performance criteria. To fill this empirical void, the present study compares six widely-used and classical classifiers while considering the ISOT Fake and True News dataset and performing the same experiments.

Table 1: Comparison of representative recent studies with the present work

Study	Main Approach	Feature/Representation			Models	Main Focus	Difference from Present Study
Recent survey [15]	Classical ML and DL	TF-IDF and representations	and	other	LR, SVM, NB, DT, RF, GB, XGBoost, DL	Comprehensive review of fake-news detection approaches	Review study rather than controlled experimental comparison
Recent empirical study [16]	ML + DL	TF-IDF, Vectorization		Count	Classical ML, RNN, BERT	Comparative evaluation of ML and DL methods	Includes deep-learning models and different experimental scope
Recent comparative study [17]	Classical ML	TF-IDF			LR, SVM, RF	Evaluation of traditional ML for fake-news detection	Uses fewer classifiers than the present study
Recent hybrid study [18]	ML + DL	TF-IDF, contextual representations	embeddings,		RF, SVM, LR, CNN, LSTM, BERT	Comparison of conventional and deep models	Focuses on hybrid/deep representations
Present study	Classical ML	TF-IDF unigrams + bigrams			LR, DT, RF, GB, PAC, XGBoost	Controlled comparison on ISOT dataset	Six classifiers evaluated under identical preprocessing, feature extraction, split, and evaluation conditions

2.2. Novelty of the Proposed Work

This Study's contribution is mostly observational and not algorithmic. This study does not present a novel machine learning classifier or a novel feature-extraction technique. Rather, it offers a controlled comparison of a handful of widely-used classical a diverse set of supervised learning models, comprising

Random Forest, XGBoost, Gradient Boosting, Logistic Regression, Decision Tree, and Passive Aggressive Classifier - applied to the same processing procedure and feature representation, namely, TF-IDF. This allows for a direct comparison of the classification performance of the different models on the ISOT Fake News Dataset, which typically consists of a point-wise comparison of the results. Using a consistent set of experimental conditions throughout the study, a direct performance comparison of different ML models on the ISOT Fake News Dataset, is typically a point-wise comparison. Besides standard performance measures, confusion matrix analysis and manual prediction examples are explored, to analyze the classification error and real behavior of the model. Therefore, the value of my research work is not laying claim to any novelty in either methodology or algorithm, but conducting a systematic empirical comparison and analysis of classic machine learning approaches while providing them with a common experimental setup.

2.3. Research Contribution and Scope

However, the present study has not evaluated transformer-based or other deep learning models, which have shown to be effective for detecting fake news. This work aimed at making a controlled empirical comparison of machine learning classifiers in comparison, based on the same TF-IDF representation. This scope was kept in practice as a computational advantage because of classical text-classification approaches that can lead to comparatively lightweight alternatives in resource-limited situations [18], [19]. The present study does, however, not state the experimentally proven superiority of the selected classical models in terms of training time, memory consumptive, model size or inference time. The computational properties of these depend on the implementation, hardware setup, feature dimensionality and the parameters of the model. A systematic evaluation of the training time, inference time, memory usage and model size is then regarded as an important direction for further experimental test. Finally, transformer-based models will be included in future research for a more comprehensive accuracy-efficiency trade-off.

Originality of the present research is, not its new algorithm, but its design and consistency of the empirical evaluation. The same ISOT dataset, identical text pre-processing, the same unigram-bigram TF-IDF representation, a fixed train-test protocol and common evaluation metrics are used for all six classifiers. This is a regulated environment which minimizes methodological variation leading to variations in results from different studies. The study also features aggregate performance figures, working with a confusion-matrix analysis and manual prediction examples to enable a comparison of the observed strengths and errors of the classifiers in the same experimental setting. The results are therefore simply an empirical yardstick against which one can compare classical machine learning methods using the chosen dataset and are not meant to assert state-of-the-art results nor methodological novelty.

3. Exploratory Data Analysis

In this study, the experiments are conducted on the ISOT fake news dataset, which is widely used in the field of automated fake news detection. The dataset has been created by the Information Security and Object Technology (ISOT) Research Laboratory at the University of Victoria, Canada. It's available in two different CSV files: Fake.csv and True.csv. Articles were collected within two categories: fake news and real news; the fake collected from sites marked as either fake or misleading, while the real news mostly consisted of articles from international news service Reuters. The dataset comes with a well-structured and relatively well-balanced set of fake news and real news articles and has therefore been widely applied towards fake news classification research.

Although the ISOT dataset is widely used as a benchmark for fake news detection, previous studies have also highlighted the need for evaluating detection models on different datasets to evaluate their ability to be generalizable across news domains as in the case of FakeNewsNet [1].

The entire data set consists of 44,898 news articles (43 of which are fake and 21,417 are authentic). There are four fields on each original article record, text, subject and date. During the preparation of the dataset, an additional label field was created to specify the label for the news type, with 0 a label of fake news and 1 a label of real news. Table 2 shows a summary of the two source files and the combined dataset produced, and Table 3 shows the distribution of the labels assigned to the combined dataset.

Table 2: Dataset description

Dataset Records	Number of Records	Columns
True Records	21,417	text, title, subject, date
Fake Records	23,481	text, title, subject, date
Combined Recors	44,898	text, title, subject, date, label

Table 3: Label distribution

News Type	Label	Articles Count
Real	1	21,417
Fake	0	23,481
Total News	-	44,898

3.1. Dataset Quality Analysis

The data set was analyzed for suitability before using the machine learning models in subsequent analysis. The fields provided such as title, text, subject, date and label were checked to see if they contained any missing values. There were no nulls found in the source data. As part of the data preparation step, however, the fillna("") function was applied to make sure empty cells do not disrupt making up of text cells or the feature extraction.

The pattern of distribution of both news categories is fairly close. Around 52.3% of the full dataset consists of fake news and around 47.7% of the full dataset consists of real news. This distribution has a near-balanced class composition and the chances of large class imbalance influence the model learning process are low. The two categories' proportion is shown in Figure 1.

The entire data set was also shuffled randomly with random_state=42 prior to creating the training and testing subsets. This will minimize the impact of the original order of the records and make it possible to repeat this scramble in later experiments. The data preparation process thus ensures a uniform randomized input to be used in the evaluation and assessment of the models and minimizes the risk that the rank order of the original records has an effect on the results reported.

3.2. Dataset Characteristics

The length of the news articles is quite different, from a short report to a long news article. In order to get the headline information, pre-processing was done on the title + text fields, combining them into one feature. This joint representation then includes more contextual information than if just headlines or article text were used separately, enabling the machine learning models to learn from both concise headlines and detailed article text.

The ISOT dataset is accepted as a benchmark dataset, but there are some limitations. Previous studies have found that many real news articles are Reuters based, and that such articles may have publisher-specific formatting (e.g., datelines and writing styles) which can make classification a relatively easy task. Thus, the high classification accuracy reported in this study needs to be viewed on the basis of this benchmark data set and further testing on external data sets is recommended for testing real-world generalization.

The distribution of the data in the class is shown in figure 1. Fake news articles make up 52.3% (23,481) while real news articles make up 47.7% (21,417). Thanks to the relatively balanced distribution, the risk of classification bias towards the majority class is minimized and a solid foundation for supervised learning is provided.

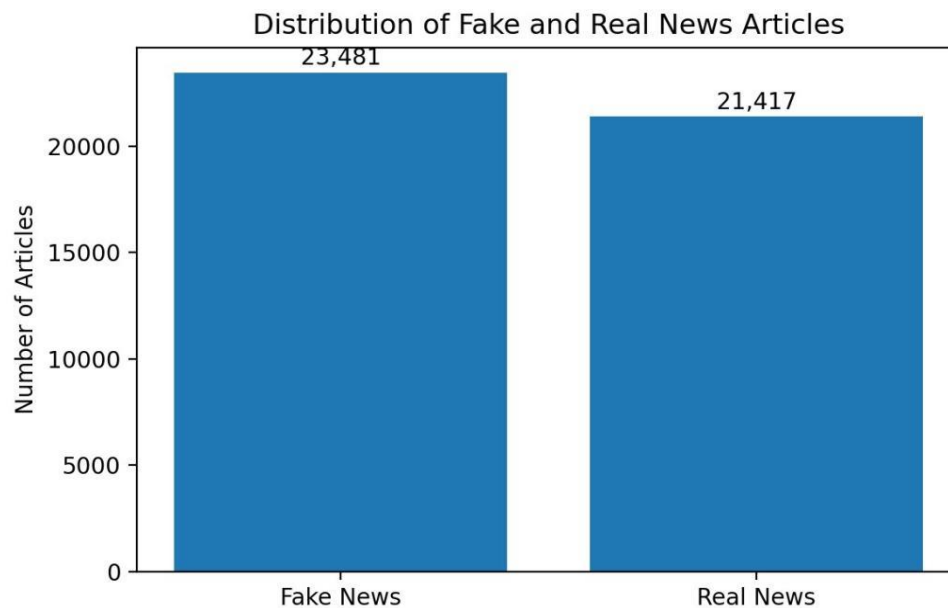


Figure 1: Distribution of fake and real news articles in the dataset

3.3. Text Length Analysis

In exploratory analysis, the lengths of news titles and article bodies were also explored. Generally, real news articles have more lengthy and detailed textual content than fake news articles, but there is a lot of overlap between the two. This observation led to the choice of the single representation of the title plus text of the article, to allow for the machine learning models to take advantage of concise information in the headline and detailed information in the article's content when the classification was made.

4. Preprocessing, Vectorization, and Feature Extraction: Fix spelling and grammar

Spelling, punctuation, grammar, capitalization, word choice, sentence structure and consistency are all checked. Identify and correct errors without altering the meaning or tone of the text or its formatting. When needed, enhance clarity and readability without needlessly altering the writer's style.

Raw news articles may include irrelevant or "noisy" information that can negatively impact the performance of machine learning, including HTML tags, URLs, punctuation, numbers and inconsistent letter casing. A full processing chain was thus put into place prior to feature extraction, and text quality was enhanced to provide for a more uniform input to all classification models.

4.1. Text Combination

The title and text columns have been combined into one column called content. Using the headline in conjunction with the article body offers a fuller representation of the text, as the headline is a summary and the body gives detailed contextual information. This is a merged representation that can be used to teach the classifiers both representations at once.

4.2. Text Cleaning

The text was subjected to a series of pre-processing steps to filter irrelevant data and normalize the text.

4.2.1. Lowercase Conversion

To make words like News, NEWS and news one and the same feature, all text was lowered to lower case. This eliminates duplication of features and reduces the dimensionality of the feature space.

4.2.2. Removal of URLs

Hyperlinks were removed as they tend not to add any substantial semantic context to classify fake news and can add noise to the data.

4.2.3. Removes HTML tags

Only the text parts of the news articles were kept and HTML elements and formatting tags were removed.

4.2.4. Removal of Punctuation and Special Characters

Non-alphabetic characters, special symbols and punctuation marks were eliminated as they seldom participate in the semantics of the text for TF-IDF based text classification and increase vocabulary size unnecessarily.

4.2.5. Removal of Numerical Values

The numerical values were not used in this dataset, because they are often not a good indicator for fake and real news.

4.2.6. Whitespace Normalization

Multiple consecutive spaces were replaced with a single space to improve text consistency and simplify tokenization. The cleaned text was saved in a new column called `clean_content` which would later be used for feature extraction. These pre-processing steps mitigate noise in the data, increase consistency and quality in the vocabulary and increase the quality of the extracted feature.

4.3. TF-IDF Feature Extraction

The text data after preprocessing was further converted into numerical vectors for input into the machine learning classifiers. The pre-processed news text was then transformed into numerical representations for classification using machine learning algorithms. The term frequency-inverse document frequency (TF-IDF) was used to derive numerical values for the terms based on their relevance in each of the documents and the collection, as a whole.

The TF-IDF representation assigns more importance to those terms that are common in a given article but rare in the complete dataset. On the other hand, words found in many documents are assigned lower weights. This representation allows the classifiers to focus further on terms giving informative information on separating fake news articles and real news articles.

The vectorizer used was tf-idf vectorizer with the following parameters:

- **stop_words = 'english'**: common English stop words (the, is, of, etc.) have been removed due to less discriminative information.
- **max_df = 0.8**: Words that occur in more than 80% of all the documents were excluded as they do not give any discrimination of the classes.
- That extremely rare words will be noisy, such those below the minimum value **min_df = 2** were omitted.
- The model included both single words (unigrams) and consecutive word pairs (bigrams) using the **ngram_range = (1,2)** parameter in the ngram model.
- **max_features = 50,000**: The number of vocabulary features used defines the size of the vocabulary. The wordnet maximum features value was set to limit the size of vocabulary used to ensure computational efficiency while maintaining the performance of the model.

- *sublinear_tf = True*: Logarithmic scaling was applied to term frequencies to reduce the influence of excessively frequent words.

4.4. Train-Test Split

The obtained dataset was split randomly into 80% for training and 20% for testing. The dataset was preprocessed and features prepared, then randomly split into 80 % for training and 20 % for testing. This left behind a training portion of 35,918 news articles, while 8,980 were held back for testing. Classifier training was conducted solely on the training set and the testing set was reserved for testing the effectiveness of classifiers on unseen articles. This separation is useful for studying the generalization performance of the model-trained classifiers.

4.5. Feature Matrix Generation

Transformer the processed articles into high dimensional sparse feature vectors that featured weight unigram and bigram features. For this training the feature matrix was of dimension 35,918 x 50,000 whereas for testing the feature matrix was of dimension, 8,980 x 50,000.

All 6 machine learning classifiers trained/tested in this study received the same training and test feature and matrix. The same feature representations were provided to each classifier class; thus, the performance could be compared under the same feature-extraction conditions.

The TF-IDF is still an approach to text representation, which is widely adopted for traditional text classification, since it assigns higher weight to informative words, and lower weight to frequently occurring words. Furthermore, it is widely used in document classification and thus can serve as a baseline representation for the tasks of detecting fake news. [2]

4.6. Cross-Validation Strategy

To minimize the effect of a single random split on the performance estimate of a model, besides the 80:20 split of train and test data, the 10-folds cross validation is often recommended. The focus of this study, however, was to compare the selected machine learning algorithms with the same experimental conditions and under the same experimental setup which was applied to test the algorithms using the hold-out validation strategy. An 80:20 train-test split was thus chosen with a fixed random state, to ensure the reproducibility and to put the results in comparison with earlier work. In future, to ensure more reliable performance estimates and checking the stability of the suggested framework, k-fold cross-validation will be implemented.

5. Model Selection and Proposed Framework

The suggested model compares multiple classifiers from the supervised classification of machine learning methods for automatic fake news recognition. The following six models were chosen for comparative analysis: Logistic Regression (LR), Decision Tree (DT), Random Forest (RF), Gradient Boosting (GB), Passive Aggressive Classifier (PAC), eXtreme Gradient Boosting (XGBoost). The classifiers are various approaches to learning that had been applied in the text classification and fake news detection studies. Applying them to the same pre-processing methods and then representing them in terms of TF-IDF features, makes it possible to compare their classification performance in different learning strategies.

In the framework one considers both linear and tree approaches under a common experimental setting. Instead of testing each classifier under various processing conditions, the selected classifiers all use the same prepared data and representation. This way, they can provide a comparison of their results with equal conditions and give an indicator as to the acceptance of their results for classification of fake news using the TF-IDF framework.

These are also very similar to those models employed in previous text classification works using sparse feature representation like TF-IDF. Some studies have focused on various classification methods for

identifying fake news: linear classification methods, algorithms based on decision trees, combination of classifiers inspired by boosting [15, [17].

5.1. Justification for Model Selection

5.1.1. Logistic Regression

Logistic Regression (LR) was added as a classic linear classification model and baseline model. It can be used in high dimensional sparse feature spaces, like those generated by TF-IDF; creates linear decision boundaries between classes. It is also relatively easy to work with and to interpret, and can be used as a benchmark to test more sophisticated classification models.

5.1.2. Decision Tree (DT)

The Decision Tree (DT) classifier employs a recursive partitioning technique in order to create decision regions of the feature space. It can model non-linear relationships between features, unlike linear models which needs feature scaling. This study uses the Decision Tree as a single tree model to compare with multiple trees models in the ensemble methods. The inclusion of such a component thus enables an understanding of the effectiveness of an individual tree model in addition to ensemble-based ones.

5.1.3. Random Forest (RF)

Random Forest (RF) is an ensemble method of multiple decision trees using bootstrap aggregated (bagging) method. The model will use the results from multiple trees in an ensemble rather than a single tree prediction. This can help minimize the variance and hence generalize better than one Decision Tree. A Random Forest machine was thus used to assess the performance of a Bagging based ensemble approach on the high dimensional feature representation.

5.1.4. Gradient Boosting (GB)

Gradient Boosting (GB) trains an ‘ensemble’ of ‘weak learners step-by-step. In the process, the following learner will take into account the mistakes made by the previous learner. By iterative nature, the method allows to represent more complex relationships between features, and offers a boosting based approach in comparison to other classifiers selected.

5.1.5. Passive Aggressive Classifier (PAC)

Passive Aggressive Classifier (PAC): A linear classification algorithm developed for the large-scale classification problem. Parameters of the classifier change during mistakes, else they are preserved. It is appropriate for high dimensional sparse vector representations like TF-IDF. Thus, a linear approach for document classification and detection of fake news was included in PAC.

5.1.6. XGBoost

EXtreme Gradient Boosting (XGBoost) is an ensemble method based on the gradient boosting method, which is further enhanced with several characteristics like regularization, parallel processing, and efficient tree construction. Its key properties are designed to enhance prediction accuracy and to prevent overfitting. For text classification tasks, such as fake news detection, XGBoost is suitable for evaluation due to its capabilities of dealing with complex interactions among features.

5.2. Rationale for Excluding Naïve Bayes and SVM

The experimental comparison did not include Naïve Bayes nor Support Vector Machine (SVM). Although both algorithms have been used extensively for fake news detection, in the present study a representative subset of classifiers spanning the linear, tree-based, ensemble and boosting approaches was

used. The models chosen are, therefore; covering the learning strategies targeted by this comparative analysis.

The reason why Naïve Bayes and SVM weren't included in the model comparison was that it wasn't the most general to detect fake news, it was due to scope of the model comparison done. Final classifiers set was chosen to have a broad interpretation of the specified machine learning paradigms satisfying moderate computation demand.

5.3. Hyperparameter Selection

The most important hyperparameters of each classifier are shown in Table 3. The values for the parameters were chosen from previous studies that had also suggested their values and from the default settings of the respective machine learning libraries. There was no extensive hyperparameter optimization done - Grid Search or Random Search. Rather, the selected values were selected for stable performance yet, letting all the analyzed classifiers be compared fairly. Further research will look into applying auto-automated Hyperparameter Optimization techniques to further improve the classification performance.

Table 4: Machine Learning Models and Hyperparameters

Model	Hyperparameters
LR	max_iter = 1000
DT	random_state = 42
GB	random_state = 42
RF	n_estimators = 100, random_state = 42, n_jobs = -1
PAC	max_iter = 1000, random_state = 42
XGBoost	n_estimators = 150, learning_rate = 0.1, max_depth = 6, eval_metric = logloss, random_state = 42

5.4. Proposed Framework

In the proposed fake news detection framework, the framework is depicted in Figure 2 that describes the overall workflow and how the planned steps are taken. First of all the Fake.csv and True.csv datasets are loaded, then the data is merged together and labels are assigned. Both the datasets are processed in a similar fashion as part of the text preprocessing as follows: all the text is converted to lower case and removal of URLs, html tags, and punctuation, numbers, and extra whitespaces. The TFIDF vectorizer (unigram, bigram features) is then used to convert the cleaned text to numerical feature vectors.

The feature vectors are split in 80:20 ratio for training and testing. The machine learning models are trained independently of each other and tested on the test set. The accuracy, precision, recall, F1-score and confusion matrix analysis are used to evaluate performance. Finally, the classification performance of best-performing classifiers is assessed and the classifier with the best performance is selected to manually classify unseen news articles.

6. Metric Selection

Five widely used classification metrics were used to test the performance of the proposed fake news detection framework: Accuracy, Precision, Recall, F1-score, and Confusion Matrix. Because fake news detection is a binary classification problem, using a single evaluation metric might not be sufficient to make a comprehensive evaluation of the performance of the model. Hence, several complementary metrics were employed to validate the classifiers in a variety of aspects.

6.1. Confusion Matrix

The Confusion Matrix is used to compile classification results by matching up labels against actual class labels. It has 4 parts:

- True Positive (TP): Fake news articles correctly identified as fake.
- True Negative (TN): Correctly identified news articles as news.
- False Positive (FP): Correct articles mistakenly identified as fake.
- False Positive (FP): Breaking news articles incorrectly identified as fake news articles.

The confusion matrix gives detailed information of the kind of errors committed by each model. It is used to detect if a classifier will tend to misclassify or not misclassify fake news articles.

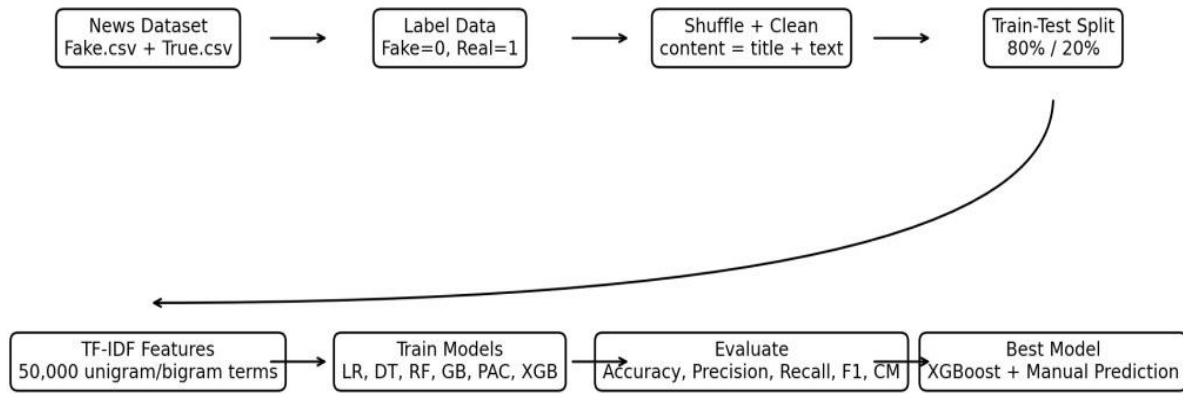


Figure 2: Proposed workflow of the fake news detection system

6.2. Accuracy

Accuracy represents the total percentage of correctly classified news articles for all the samples in test set. If the accuracy of the classifier is high, it means that it is easily able to differentiate among fake news and real news articles. But even its accuracy might not fully amount to performance when the classes do not have equal distributions.

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

6.3. Precision

Precision is the percentage of news articles correctly predicted as fake.

$$\text{Precision} = \frac{TP}{TP+FP} \quad (2)$$

A high precision value means that there are relatively few false positive predictions by the classifier. High precision results in the least risk of mislabeling news as fake in fake news detection.

6.4. Recall

The percentage of true fake news documents which are identified as fake news.

$$\text{Recall} = \frac{TP}{TP+FN} \quad (3)$$

High recall scores mean that the classifier can catch the majority of the fake news articles and missed detection rate is low. This is a significant measure because spreading undetected fake news could lead to the propagation of misinformation.

6.5. F1-Score

The F1-score is the harmonic mean of precision and recall.

$$F1 - score = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (4)$$

The F1-score gives a balance evaluation because both the false positive and false negative are taken into consideration. It is particularly useful for comparing different classifiers with varying precision / recall compromise.

6.6. Importance of Multiple Evaluation Metrics

The different evaluation metrics measure different aspects of the classifiers' performance. The accuracy measure provides an overall evaluation of the accuracy of the predictions, precision is used for measuring prediction reliability, recall for measuring the ability to detect fake news, and the F1-score for balancing both measures. Confusion matrix provides additional details on how many correct and incorrect predictions have been made. Hence, all the five metrics were used simultaneously for a clear and sound comparison of performances of six machine learning models evaluated in this work in the Table 4.

7. Results and Discussion

TF-IDF feature extraction then resulted in the creation of training and testing datasets, which were then used to evaluate the performance of six ML models. Ensuring a fair comparison between all classifiers, the same processing was applied to the data to form classifiers, and the same test set was used for evaluation, with accuracy, precision, recall, F1-score, and confusion matrix analysis performed on the test set. Table 5 shows the quantitative performance of each of the classifiers.

Table 5: . Performance comparison of machine learning models

Machine Learning Algorithm	Training Accuracy	Testing Accuracy	Precision	Recall	F1-Score
Logistic Regression	0.9963	0.9923	0.9889	0.9951	0.9920
Decision Tree	1.0000	0.9960	0.9967	0.9949	0.9958
Random Forest	1.0000	0.9957	0.9937	0.9972	0.9955
Gradient Boosting	0.9977	0.9962	0.9942	0.9979	0.9960
Passive Aggressive Classifier	1.0000	0.9972	0.9958	0.9984	0.9971
XGBoost	0.9999	0.9986	0.9979	0.9991	0.9985

7.1. Comparative Evaluation of Machine Learning Models

Table 5 indicates that all of the six classifiers exhibited high testing accuracy (more than 99%), suggesting that the TF-IDF features offer highly discriminate representations for ISOT dataset. The best performing models with highest testing accuracy were XGBoost (99.86 %), Passive Aggressive Classifier (99.72 %), Gradient Boosting (99.62 %), Decision Tree (99.60 %), Random Forest (99.57 %) and Logistic Regression (99.23 %). This result is similar to a previous comparative study, which reported the effectiveness of boosting algorithms under development to model complex feature interactions along with the regularization of their overfitting through regularization mechanisms, followed by the Passive Aggressive Classifier (99.72%), the Gradient Boosting (99.62%), the Decision Tree (99.60%), the Random

Forest (99.57%), and the Logistic Regression (99.23%). The same kind of behavior was found in the case of precision, recall and the F1 measure, which in this case also showed the highest value for XGBoost.

7.2. Discussion of Experimental Results

Focusing on the ensemble boosting strategy of XGBoost gives us the reason for this superior performance. XGBoost is an extension of the SNOWFALL algorithm which builds several trees sequentially, each one rectifying the previous trees' mistakes. Additionally, XGBoost models come with regularization methods to prevent overfitting and efficiently capture complex relationships among TF-IDF features. The model achieved these features that improve the classification of fake and genuine news articles than the other classifiers assessed.

Excellent performance was also outlined for the Passive Aggressive Classifier, which a slight peculiarity is tailored for high-dimensional sparse data sets like TF-IDF feature matrices. Given that the strategy of online learning only updates model parameters when prediction errors are incurred, it is quite efficient and also highly effective for text classification problems.

The lowest testing accuracy was achieved using logistic Regression using among the models included in the evaluation. It is good at covering the simple linear decision boundary, and functions well when there are few textual features; however, it is not a good baseline for the complex nonlinear relations of textual features. In a similar manner, the nonlinear pattern modelled by Decision Tree and Random Forest showed some inferiority to the boosting-based patterns modelled above, in this study.

The outcomes overall show that the ensemble boosting methods along with TF-IDF feature extraction techniques are helpful to enhance the predictive performance for detecting fake news.

7.3. Confusion Matrix Analysis

To gain more insight into the classification's performance of each model the confusion matrices are included in Table 6. Of all the fake news samples (4687) and authentic news samples (4280) given in the test set, XGBoost had only 13 samples that were misclassified. It is the smallest number of classification errors which has been observed with all models being assessed. The Passive Aggressive Classifier and Logistic Regression had 25 and 69 misclassifications, respectively, compared to the DRL, which had 24. The results show that XGBoost shows the most well-balanced classification results among all news categories, both fake and real. Thus, very little should be taken from the current results as an indication of the most efficient performance of the highest inductive complexity model; instead, this should be treated as primarily a comparison of performance with the predictor.

Table 6: Summary of Confusion Matrices

Model	True Fake	Fake Predicted Real	Real Predicted Fake	True Real	Total Errors
Logistic Regression	4,648	48	21	4,263	69
Decision Tree	4,682	14	22	4,262	36
Random Forest	4,669	27	12	4,272	39
Gradient Boosting	4,671	25	9	4,275	34
Passive Aggressive Classifier	4,678	18	7	4,277	25

XGBoost	4,687	9	4	4,280	13
----------------	-------	---	---	-------	----

7.4. Interpretation of the High Accuracy

These results are very highly accurate based on the classification obtained; however, they should be taken with a grain of salt. Experiments were performed with the popular ISOT Fake News Dataset (IDFD), which can be used as a standard test set for the study of fake news detection. Studies have labeled several real news articles from Reuters as publisher-specific information, including formatting, writing style, and dateline information, though, previous studies have reported that many real news articles come from Reuters. As a result, misinformation related linguistic patterns may not be the only feature that machine learning models are able to capture. Hence the accuracy reported must be considered in reference to the ISOT dataset and not be interpreted as an accuracy in a real application. Further benchmarks measurements need to be further assessed on other datasets and preprocessing approaches investigated which might allow some publisher-specific artifacts to be removed prior to training.

7.5. Statistical Interpretation

While the classifier which performed the best on the test set was XGBoost, with several classifiers having only a slight difference in performance. So, the winner of XGBoost shouldn't be taken as a statistical winner, but as the top performer under this experimental setup. In the coming future, a more demanding comparison will be performed by k-fold cross validation and the statistical significance test of McNemar, to identify if the performance differences observed are statistically significant.

While the statistical significance testing methods like McNemar's test testing competing classifiers would be useful, they were not in the scope of the present study. Instead, multiple evaluation metrics such as accuracy, precision, recall, F1-score, and confusion matrix analysis were utilized for modeling in order to compare the models. Future work will involve using statistical hypothesis testing methods to see if the differences among the results of the classifiers is statistically significant.

7.6. Computational Considerations

An important practical consideration in choosing an appropriate fake news detection model, is its computational efficiency. When computation is limited, conventional machine learning methods based on sparse TF-IDF representations could offer relatively lightweight alternatives to high-capacity deep learning architectures, especially in applications such as these [18, 19]. But there is more to it than just the type of classifier—training and inference costs are influenced by the number of features, the size of the data set, implementation details, hardware configuration and model parameters.

The evaluation was focused mainly on the prediction, but not on direct computational benchmarking in the present study. For this reason, training time, inference time, memory usage, model size and hardware usage were not used as outcome measures in the experiments and no real efficiency improvement is provided. Such high predictive accuracy for XGBoost should thus be understood as a measure of the success of the XGBoost's classification when tested against the standardized data set. For the future, a controlled computational benchmark of all the classifiers evaluated will be performed with the same hardware and software settings to quantify their accuracy–efficiency trade-offs.

7.7. Manual Prediction Analysis

To assess the feasibility of the proposed model, a manual prediction experiment was conducted with the news text inputs by the users. Although it sounds like an actual news statement, the example in Table 7 was tagged as Fake News. This outcome highlights an important constraint that is inherent in the proposed model. The model was built mostly from news articles and not individual sentences, so this can lead to the classification of statements being short and context-limited. The same can be done by further testing of the framework with more manually curated illustrative cases belonging to different topic-based settings and writing styles, in order to offer more effective real-world testing of the framework.

As shown by the manual predicate prediction examples, the proposed system is performed well for complete news statements with enough contextual information. But short or ambiguous statements may be misclassified, they don't have the linguistic patterns in the training data. This observation highlights the need in analyzing fake news detection models with different real-world news articles and various-length datasets.

Table 7: Sample manual prediction result

Sample	Expected Label	Predicted Label
Government announces new education policy for schools.	Real	Fake
Scientists discover a new treatment for diabetes after clinical trials.	Real	Real
Celebrity secretly owns a spaceship on the moon.	Fake	Fake
Earth will become dark for 10 days next month according to NASA.	Fake	Fake
Parliament approves the annual national budget after debate.	Real	Real

8. Conclusion

The aim of this research was to compare the performances of six classical machine learning classifiers in an empirical way to detect fake news using textual features extracted from the TF-IDF approach. All evaluated models had high classification capabilities on the ISOT Fake and True News Dataset under the same preprocessing conditions, feature extraction and evaluation. XGBoost showed the best ever achieved results with its testing accuracy of 99.86%, its precision of 99.79%, its recall of 99.91%, its F1-score of 99.85%, and with only 13 classification errors.

The results show the potential of classical machine learning algorithms and TF-IDF to achieve extremely good results in the classification of fake news from the chosen benchmark dataset. The results of this research should be viewed in the context of the limitations of the ISOT dataset and the experimental design, as well as in relation to the generalization to real-world scenarios, and should not be interpreted as evidence of the superiority of models, or as evidence of generalization to real-world scenarios.

Externally testing models on relevant data sets and running programmed comparisons with transformer-based methods and efficiency measurements concerning predictive performance and utility are suggested for future work to draw a broader evaluation of predicts and real-life applicability.

9. Limitations

While the framework proposed above performed well for the ISOT data set, some limitations should be noted.

- One benchmark dataset has been used for the experiments. Thus, the generalizability of the model to news articles published by other news publishers, domains or date may not be fully captured in the reported performance.
- The data source used in this paper includes mostly news gathered years earlier. Misinformation is often constantly changing, which can affect the linguistic patterns that can be detected in historical data and those found in current online news and social media sites.
- The other drawback of this study is that the framework was tested on ISOT Fake News Dataset only. Despite its utility in fake news research, more evaluation on other benchmark datasets, such as LIAR and FakeNewsNet, would enable more a robust evaluation of model generalization.

Considering the time and computational limitations: external dataset validation was not conducted in the present study, and is recommended as a direction for future studies.

- The train-test split of 80:20 was used in this study, without k-fold cross validation. This method is frequently adopted in the machine learning research community, but future research should include multiple cross validations and statistical significance testing to employ more robust performance estimation.
- Last, the suggested framework doesn't go so far as to verifying facts, but it does do text classification. Therefore, the model relies on learned patterns of language to assign the label to them instead, and will not independently test whether the factual statements in a news article are actually true or false.

10. Future Work

A few directions are here which further improve the proposed framework:

- **External Dataset Validation:** Usage of other benchmark datasets such as LIAR and FakeNewsNet could be done in future studies to test the proposed model's ability to generalize beyond the ISOT dataset.
- **Transformer-Based Language Models:** Natural language understanding has been shown to be extraordinary by recent transformer models such as BERT, RoBERTa and FakeBERT. These deep learning models could be compared with the classical machine learning methods developed as part of this study in future research.
- Recent work has shown that BERT-based models, like transformer networks, greatly outperform a number of traditional machine learning models in the context of better contextual representation learning and can be used to detect fake news more accurately than many other traditional models of faking news detection [3].
- **Hyperparameter Optimization:** Automated hyperparameter optimization techniques such as Grid Search or Random Search or Bayesian Optimization might also be used to enhance the performance of the classifiers, and to identify the optimum values for a hyperparameter.
- **Explainable Artificial Intelligence (XAI):** To enhance the interpretability of model predictions, further research could integrate explainable AI techniques such as SHAP or LIME and boost the trust in automatic fake news detection systems by the public.
- **Real-Time Fake News Detection:** The proposed framework can be adapted to be synthesized into an online real-time web application that can analyze the online news articles and social media content as they are being published, to support timely detection of misinformation.

Funding Statement: The author received no specific funding for this study.

Declaration of AI Use: AI tools were used for manuscript language enhancement, while the authors retained full responsibility for the content.

Conflicts of Interest: The authors declare no conflict of interest regarding the publication of this study.

Data Availability: The dataset used in this study is the publicly available ISOT Fake News Dataset.

References

- [1] Shu, K., Mahudeswaran, D., Wang, S., Lee, D., & Liu, H. (2020). FakeNewsNet: A Data Repository with News Content, Social Context, and Spatiotemporal Information for Studying Fake News on Social Media. *Big Data*, 8(3), 171–188. <https://doi.org/10.1089/big.2020.0062>
- [2] Khan, J. Y., Khondaker, Md. T. I., Afroz, S., Uddin, G., & Iqbal, A. (2021). A benchmark study of machine learning models for online fake news detection. *Machine Learning with Applications*, 4, 100032. <https://doi.org/10.1016/j.mlwa.2021.100032>

- [3] Kaliyar, R. K., Goswami, A., & Narang, P. (2021). FakeBERT: Fake news detection in social media with a BERT-based deep learning approach. *Multimedia Tools and Applications*, 80(8), 11765–11788. <https://doi.org/10.1007/s11042-020-10183-2>
- [4] Jouhar, J., Pratap, A., Tijo, N., & Mony, M. (2024). Fake News Detection using Python and Machine Learning. *Procedia Computer Science*, 233, 763–771. <https://doi.org/10.1016/j.procs.2024.03.265>
- [5] Sharma, U., Saran, S., & Patil, S. M. (2020). Fake news detection using machine learning algorithms. *International Journal of creative research thoughts (IJCRT)*, 8(6), 509-518.
- [6] Khanam, Z., Alwasel, B. N., Sirafi, H., & Rashid, M. (2021). Fake News Detection Using Machine Learning Approaches. *IOP Conference Series: Materials Science and Engineering*, 1099(1), 12040. <https://doi.org/10.1088/1757-899x/1099/1/012040>
- [7] Ahmed, H., Traore, I., & Saad, S. (2017). Detecting opinion spams and fake news using text classification. *Security and Privacy*, 1(1). Portico. <https://doi.org/10.1002/spy2.9>
- [8] Ahmed, H., Traore, I., & Saad, S. (2017). Detection of Online Fake News Using N-Gram Analysis and Machine Learning Techniques. *Intelligent, Secure, and Dependable Systems in Distributed and Cloud Environments*, 127–138. https://doi.org/10.1007/978-3-319-69155-8_9
- [9] Ahmad, I., Yousaf, M., Yousaf, S., & Ahmad, M. O. (2020). Fake News Detection Using Machine Learning Ensemble Methods. *Complexity*, 2020, 1–11. <https://doi.org/10.1155/2020/8885861>
- [10] Reis, J. C. S., Correia, A., Murai, F., Veloso, A., & Benevenuto, F. (2019). Explainable Machine Learning for Fake News Detection. *Proceedings of the 10th ACM Conference on Web Science*, 17–26. <https://doi.org/10.1145/3292522.3326027>
- [11] Kumar, S., Asthana, R., Upadhyay, S., Upreti, N., & Akbar, M. (2019). Fake news detection using deep learning models: A novel approach. *Transactions on Emerging Telecommunications Technologies*, 31(2). Portico. <https://doi.org/10.1002/ett.3767>
- [12] Wang, Y., Yang, W., Ma, F., Xu, J., Zhong, B., Deng, Q., & Gao, J. (2020). Weak Supervision for Fake News Detection via Reinforcement Learning. *Proceedings of the AAAI Conference on Artificial Intelligence*, 34(01), 516–523. <https://doi.org/10.1609/aaai.v34i01.5389>
- [13] Ibrishimova, M. D., & Li, K. F. (2019). A Machine Learning Approach to Fake News Detection Using Knowledge Verification and Natural Language Processing. *Advances in Intelligent Networking and Collaborative Systems*, 223–234. https://doi.org/10.1007/978-3-030-29035-1_22
- [14] Nasir, J. A., Khan, O. S., & Varlamis, I. (2021). Fake news detection: A hybrid CNN-RNN based deep learning approach. *International Journal of Information Management Data Insights*, 1(1), 100007. <https://doi.org/10.1016/j.jjime.2020.100007>
- [15] Alghamdi, J., Luo, S., & Lin, Y. (2023). A comprehensive survey on machine learning approaches for fake news detection. *Multimedia Tools and Applications*, 83(17), 51009–51067. <https://doi.org/10.1007/s11042-023-17470-8>
- [16] Alghamdi, J., Lin, Y., & Luo, S. (2022). A Comparative Study of Machine Learning and Deep Learning Techniques for Fake News Detection. *Information*, 13(12), 576. <https://doi.org/10.3390/info13120576>
- [17] Farhangian, F., Cruz, R. M. O., & Cavalcanti, G. D. C. (2024). Fake news detection: Taxonomy and comparative study. *Information Fusion*, 103, 102140. <https://doi.org/10.1016/j.inffus.2023.102140>
- [18] Hamed, S. K., Ab Aziz, M. J., & Yaakub, M. R. (2023). A review of fake news detection approaches: A critical analysis of relevant studies and highlighting key challenges associated with the dataset, feature representation, and data fusion. *Heliyon*, 9(10), e20382. <https://doi.org/10.1016/j.heliyon.2023.e20382>
- [19] Saeed, A., & Solami, E. A. (2023). Fake News Detection Using Machine Learning and Deep Learning Methods. *Computers, Materials & Continua*, 77(2), 2079–2096. <https://doi.org/10.32604/cmc.2023.030551>