Power System Technology
ISSN:1000-3673

Received: 06-01-2026 Revised: 15-02-2026 Accepted: 12-03-2026

# A Comparative Analysis of Text Vectorization and Machine Learning Classifiers for Fake News Detection

[1] Dr. Ashutosh Dhamija, [2]Dr. Mukesh Kumar

[1]Associate Professor, Department of Electronics & Communication Engineering, Maharishi Markandeshwar Engineering College, Maharishi Markandeshwar (Deemed to be University), Mullana, Ambala, Haryana-133207, India.

ashutosh.dhamija@mmumullana.org

[2]Associate Professor, Department of Electronics & Communication Engineering, The Technological Institute of Textile and Sciences, Bhiwani, Haryana-127021, India.

drmukeshkumar@titsbhiwani.ac.in

**Abstract:** In today's digital era, the media landscape has seamlessly transitioned from print to online platforms, leading to an unprecedented increase in information accessibility and exchange. However, this transformation has also intensified a major challenge, the rapid proliferation of fake news, which refers to fabricated or misleading information that can be easily produced and disseminated. This paper addresses the growing global concern of misinformation and explores potential solutions through machine learning techniques. The proposed study develops a model designed to assess the authenticity of news articles by evaluating multiple text vectorization methods, specifically the Bag-of-Words approach using both Count Vectorizer and TF-IDF Vectorizer. Two classification algorithms, namely the Multinomial Naive Bayes Classifier and the Passive Aggressive Classifier**,** are employed to detect fake news. The study further investigates how text pre-processing influences overall model performance. The dataset chosen for training is comprised of 67.7% curated information, while the remaining 33.3% remains untrained raw data. Notably, the model demonstrates a noteworthy efficiency rate of 93.78% under optimum conditions. This strong result demonstrates how well the suggested methodology works to differentiate between real and fake news.

**Keywords:** Fake News Detection, Text Vectorization, TF-IDF, Machine learning and Feature Extraction

## 1. Introduction

In the ever-evolving landscape of today's digital era, the media sector has seamlessly transitioned from traditional print formats to digital platforms, fundamentally reshaping the dynamics of information production and consumption. However, amidst this transformative shift, a formidable challenge has emerged—the widespread proliferation of "fake news." This

Received: 06-01-2026 Revised: 15-02-2026 Accepted: 12-03-2026

term encapsulates the alarming surge in the creation, dissemination, and accessibility of misleading information within the interconnected and digitized global community.

The phenomenon of fake news has evolved into a critical societal concern, posing a tangible threat to the very foundations of informed decision-making and public discourse. In the age of instantaneous information sharing on digital platforms, erroneous narratives can swiftly gain traction, exerting influence on opinions, shaping perspectives, and even impacting socio-political landscapes. The repercussions of misinformation extend beyond individual beliefs, permeating vital sectors such as politics, healthcare, and societal trust. The urgency to confront and mitigate this challenge lies in safeguarding the integrity of information dissemination and ensuring a well-informed public discourse.

In response to the escalating prevalence of disinformation, this research endeavours to propose a comprehensive solution by focusing on the development of a robust model designed to assess the reliability of news stories. The primary objective is to forge an effective mechanism capable of distinguishing between authentic and fabricated news, thereby fortifying the digital media landscape against the pernicious effects of misinformation.

The envisaged solution involves a meticulous evaluation process, leveraging various text vectorization approaches, with a particular emphasis on the Bag of Words methodology. This encompasses the utilization of both Count Vectorizer and TF-IDF Vectorizer, complemented by the application of sophisticated classification algorithms, such as the Multinomial Naive Bayes Classifier and the Passive Aggressive Classifier. Additionally, the study systematically examines the impact of text pre-processing on the overall performance of the proposed model.

By delving into the intricacies of fake news detection and presenting a concrete and innovative solution, this research aspires to illuminate the challenges posed by misinformation, thereby paving the way for more resilient and informed digital communication. Subsequent sections will meticulously explore the methodological nuances, experimental outcomes, and broader implications of our research.

**2. Literature Review**

[1] The issue of fake news has become a growing concern in today's digital age, and detecting it has become a challenging task. This study provides an approach for identifying false news using supervised machine learning algorithms on a labeled dataset. In order to achieve the maximum level of accuracy, feature selection techniques are used to experiment and choose the best-fit features from the dataset, which is manually categorized and guaranteed. The suggested model employs many classification methods, such as Random Forest, Support Vector Machines, and Naïve Bayes. To achieve optimal accuracy and precision, trials are

Power System Technology
ISSN:1000-3673
Received: 06-01-2026 Revised: 15-02-2026 Accepted: 12-03-2026

carried out separately for each technique and in combinations with one another. The results show that the proposed model achieves high accuracy in detecting fake news, with the highest accuracy being 92%. The suggested model offers a reliable automatic index rating or scoring system for the trustworthiness of various publications and news contexts, and it may be utilized and integrated with any system for future usage. Overall, this paper provides valuable insights into the use of machine learning approaches for detecting fake news and highlights the potential applications of fake news detection in real-world scenarios.

[2] The paper discusses various approaches to detecting fake news, including linguistic analysis, unified key sentence information, and machine learning techniques such as Naïve Bayes classifier, SVM, and NLP. According to the paper, "most of the smart phone users prefer to read the news via social media over internet". This has led to an increase in the spread of fake news through social media platforms such as WhatsApp, Facebook, and Twitter. The paper emphasizes the importance of authenticating news sources and detecting fake news to prevent the spread of rumours and misinformation. This study proposes a system that aggregates news and utilizes Support Vector Machine to identify whether it is authentic or fraudulent. It does this by using machine learning and natural language processing. The suggested model's findings indicate an accuracy of up to 93.6%. The paper concludes that this system has the potential to greatly improve the accuracy and reliability of news sources and combat the spread of fake news on social media platforms.

[3] The objective of this paper is to identify the effects of fake news and to detect it using machine learning classifiers. This study presents a strategy for detecting false news that combines machine learning (ML) classifiers with a systematic literature review (SLR). The dataset used in this study is 50% false news and 50% actual news, and it is accessible to the public on Kaggle. The ML classifiers used in this study include logistic regression (LR), k-nearest neighbour (KNN), decision tree (DT), random forest (RF), and support vector machine (SVM). The proposed method achieved an accuracy of 98.5% in detecting fake news using the LR classifier, which outperformed the other classifiers. The study also found that fake news can have a significant impact on people's beliefs and behaviours, leading to negative consequences. Therefore, it is crucial to detect and prevent the spread of fake news. The proposed method can be further improved by using more advanced ML techniques and by incorporating additional features such as social engagement and writing style. Moreover, the study can be extended to other languages and countries to investigate the impact of fake news on a global scale.

[4] The paper begins with a literature review of previous research on fake news detection, including machine learning and deep learning techniques. The authors then describe their methodology, which involves using a dataset of fake and real news articles and applying data pre-processing, data exploration, and machine learning classification models to detect fake

Received: 06-01-2026 Revised: 15-02-2026 Accepted: 12-03-2026

news. The algorithms used in this study showed an accuracy of more than 99%, which is almost perfect. The authors also note that there are limitations and insufficiencies in their approach, such as unbalanced or uncleaned datasets, which can lead to inaccurate results. The paper concludes with a discussion of future work in this field, including the application of sentiment analysis and deep learning-enabled big data models. Overall, this paper provides a comprehensive analysis of fake news detection using machine learning techniques and offers valuable insights into the challenges and opportunities in this field. Technical details of the paper include the use of Spark machine learning and the application of a 4-g model with term frequency and TF-IDF to extract fake contents.

[5] The paper begins with an abstract that summarizes the key findings of the research. The authors present a statistical analysis of a dataset of online articles with real and fake classes, along with sentiment analysis and irony detection systems. They also provide insights into the related work and methodology used in the research. In the related work section, the authors discuss the existing approaches, challenges, and observations from other languages that can be applied to Romanian resources. They highlight the importance of natural language processing techniques in automated analysis of political statements. The methodology section describes the proposed method in detail, including the dataset used, feature extraction, and classification techniques. The authors also discuss the evaluation metrics used to measure the accuracy of their approach. In the experiments section, the authors present the results of their approach and compare it with other state-of-the-art methods. They report an accuracy of 91.5% in detecting fake news, which is a significant improvement over existing method. In conclusion, this paper provides valuable insights into the detection of fake news in the Romanian online news landscape. The authors' approach using statistical analysis, sentiment analysis, and irony detection systems has shown promising results in identifying fake news with high accuracy. This research can be useful for policymakers, journalists, and the general public in combating the spread of fake news.

## 3. Materials and methods

### 3.1. Proposed method

Figure 1 illustrates the proposed methods and their flow. The proposed methods primarily encompass data collection, optional data pre-processing, feature extraction, model selection, model training, and model evaluation.

Power System Technology
ISSN:1000-3673
Received: 06-01-2026 Revised: 15-02-2026 Accepted: 12-03-2026

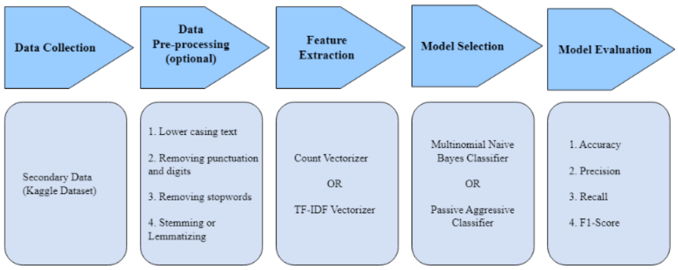


Figure 1. Proposed model diagram

3.2. Data collection

The dataset utilized in this research is publicly accessible on Kaggle, comprising a total of 6335 news articles labeled as either fake or real. Figure 2 illustrates the distribution of fake and real news in dataset.

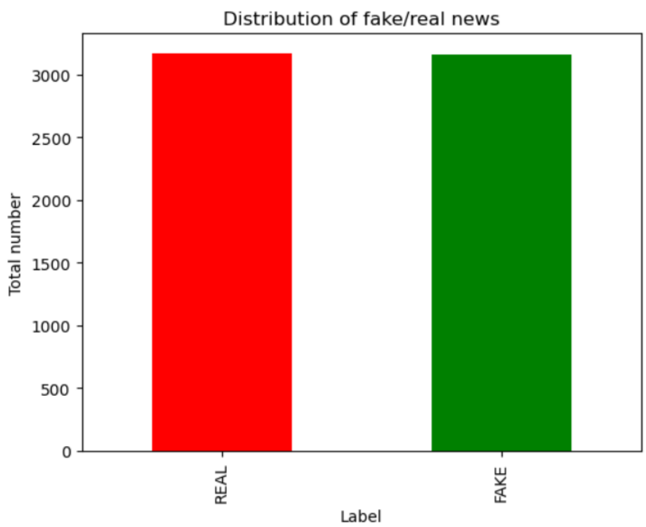


Figure 2. Distribution of fake news and real news in dataset

3.3. Data pre-processing (optional)

The raw dataset comprises news articles labeled as either fake or real. Although this step is optional, it enhances the efficiency of model training. Therefore, before initiating the training of the model, the dataset undergoes the following steps: converting the text in the dataset to lowercase, removing non-alphabetic characters from the text, eliminating stop words (which do not carry significant meaning), and stemming or lemmatizing the text (aimed at reducing the dimensionality of text data). The visual representation of the dataset before and after pre-processing is depicted in Figures 3 and 4.

Received: 06-01-2026 Revised: 15-02-2026 Accepted: 12-03-2026

| Unnamed: 0 | title | text | label |
|---|---|---|---|
| 8476 | You Can Smell Hillary's Fear | Daniel Greenfield, a Shillman Journalism Fello... | FAKE |
| 10294 | Watch The Exact Moment Paul Ryan Committed Pol... | Google Pinterest Digg Linkedin Reddit Stumbleu... | FAKE |
| 3608 | Kerry to go to Paris in gesture of sympathy | U.S. Secretary of State John F. Kerry said Mon... | REAL |
| 10142 | Bernie supporters on Twitter erupt in anger ag... | — Kaydee King (@KaydeeKing) November 9, 2016 T... | FAKE |
| 875 | The Battle of New York: Why This Primary Matters | It's primary day in New York and front-runners... | REAL |
| 6903 | Tehran, USA | \nI'm not an immigrant, but my grandparents ... | FAKE |

Figure 3. Text data before pre-processing

| Unnamed: 0 | title | text | label |
|---|---|---|---|
| 8476 | smell hillari ' fear | daniel greenfield shillman journal fellow free... | FAKE |
| 10294 | watch exact moment paul ryan commit polit suic... | googl pinterest digg linkedin reddit stumbleup... | FAKE |
| 3608 | kerri go pari gestur sympathi | us secretari state john f kerri said monday st... | REAL |
| 10142 | berni support twitter erupt anger dnc tri warn | — kayde king kaydeek novemb lesson tonight dem... | FAKE |
| 875 | battl new york primari matter | primari day new york frontrunn hillari clinton... | REAL |
| 6903 | tehran usa | ' immigr grandpar year ago arriv new york citi... | FAKE |

Figure 4. Text data after pre-processing

3.4. Feature extraction

As machine learning operates solely on numerical data, a straightforward bag-of-words approach is employed. In this method, each article is represented as a vector of word frequencies. The research utilizes the following techniques: 1. Count Vectorizer, 2. TF-IDF Vectorizer.

3.4.1. Count Vectorizer

The Count Vectorizer functions by initially tokenizing the text data into words and subsequently counting the frequency of each word in each document of Figures 5 and 6.

| | 00 | 000 | 0000 | 00000031 | 000035 | 00006 | 0001 | 0001pt | 000ft | 000km | ... | حلب | عربي | عن | لم | ما | محاولات | من | هذا | والمرضى | ɛɔade |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | ... | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | ... | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | ... | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 3 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | ... | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 4 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | ... | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

5 rows × 60400 columns

Figure 5. Applied Count Vectorizer on raw dataset

| | aa | aaa | aab | aadmi | aah | aaib | aalia | aam | aamaq | aamon | ... | حلب | عربي | عن | لم | ما | محاولات | من | هذا | والمرضى | ɛɔad |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | ... | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | ... | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | ... | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 3 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | ... | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 4 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | ... | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

5 rows × 50924 columns

Figure 6. Applied Count Vectorizer on pre-processed dataset

3.4.2. TF-IDF Vectorizer

TF-IDF, which stands for Term Frequency Inverse Document Frequency, operates similarly to the count vectorizer. However, it goes a step further by considering the frequency of a

Power System Technology
ISSN:1000-3673
Received: 06-01-2026 Revised: 15-02-2026 Accepted: 12-03-2026

word across the entire corpus, assigning a weight to each word based on its importance in the document of Figures 7 and 8.

| | 00 | 000 | 0000 | 00000031 | 000035 | 00006 | 0001 | 0001pt | 000ft | 000km | ... | حلب | عربي | عن | لم | ما | محاولات | من | هذا | والمرضى | ยงade |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | ... | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| 1 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | ... | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| 2 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | ... | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| 3 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | ... | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| 4 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | ... | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |

5 rows × 60095 columns

Figure 7. Applied TF-IDF Vectorizer on raw dataset

| | aa | aaa | aab | aadmi | aah | aaib | aalia | aam | aamaq | aamon | ... | حلب | عربي | عن | لم | ما | محاولات | من | هذا | والمرضى | ยงad |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | ... | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| 1 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | ... | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| 2 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | ... | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| 3 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | ... | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| 4 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | ... | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |

5 rows × 50682 columns

Figure 8. Applied TF-IDF Vectorizer on pre-processed dataset

### 3.5. Model selection

During this stage, classical machine learning algorithms are employed for the classification of the data.

#### 3.5.1. Multinomial Naïve Bayes Classifier

The Multinomial Naïve Bayes classifier utilizes the principles of Bayes theorem to assign documents to classes. Specifically designed for Natural Language Processing (NLP), it represents a variation of the Bayes algorithm.

##### 3.5.1.1. Count Vectorizer

As depicted in Figure 9, the confusion matrix represents the performance of the model trained using the count vectorizer and Multinomial Naïve Bayes Classifier on the raw dataset.

Power System Technology
ISSN:1000-3673
Received: 06-01-2026 Revised: 15-02-2026 Accepted: 12-03-2026

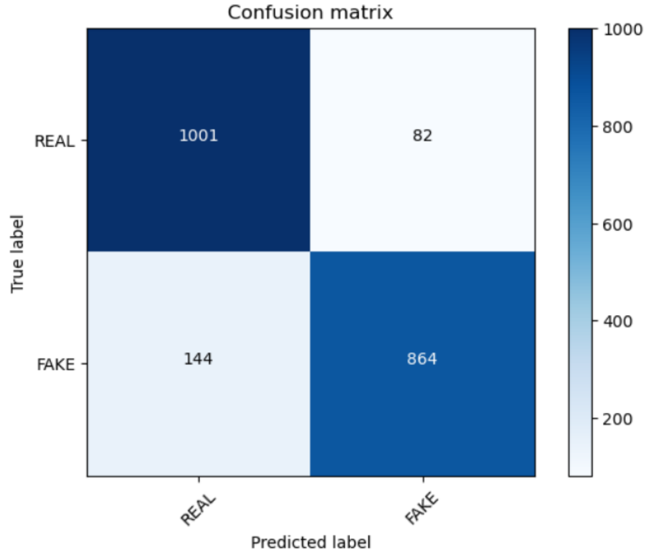


Figure 9. Count vectorizer and Multinomial Naïve Bayes Classifier (Raw dataset)

As depicted in Figure 10, the confusion matrix represents the performance of the model trained using the count vectorizer and Multinomial Naïve Bayes Classifier on the pre-processed dataset.

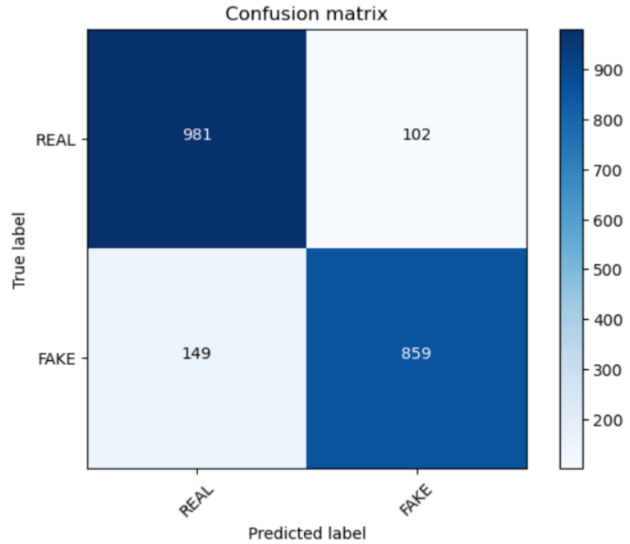


Figure 10. Count vectorizer and Multinomial Naïve Bayes Classifier (Pre-processed dataset)

3.5.1.2. TF-IDF Vectorizer

As depicted in Figure 11, the confusion matrix represents the performance of the model trained using the TF-IDF vectorizer and Multinomial Naïve Bayes Classifier on the raw dataset.

Power System Technology
ISSN:1000-3673
Received: 06-01-2026 Revised: 15-02-2026 Accepted: 12-03-2026

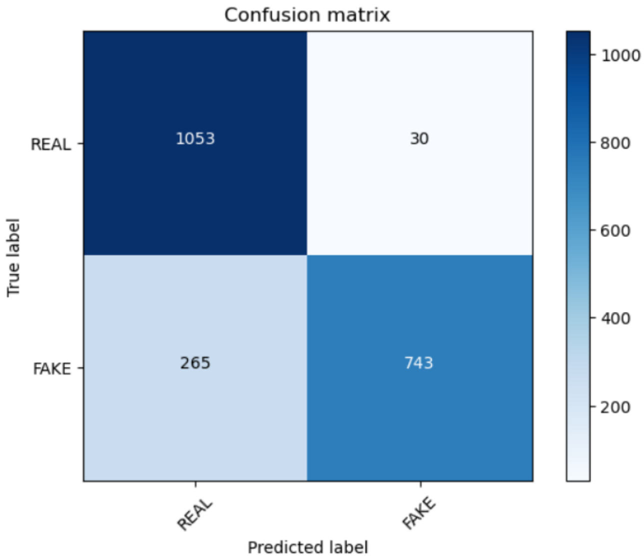


Figure 11. TF-IDF vectorizer and Multinomial Naïve Bayes Classifier (Raw dataset)

As depicted in Figure 12, the confusion matrix represents the performance of the model trained using the TF-IDF vectorizer and Multinomial Naïve Bayes Classifier on the pre-processed dataset.

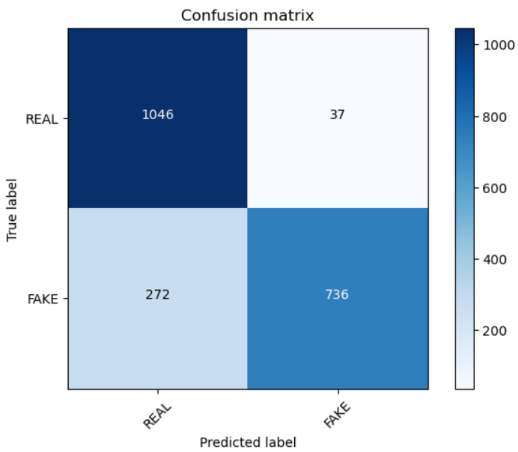


Figure 12. TF-IDF vectorizer and Multinomial Naïve Bayes Classifier (Pre-processed dataset)

3.5.2. Passive Aggressive Classifier

Passive-Aggressive Algorithms for Large-Scale Learning represent a family of machine learning algorithms tailored for in-depth data analysis. Resembling the Perceptron, these algorithms deviate by not depending on a fixed learning rate and stand out by incorporating a regularization parameter.

Power System Technology
ISSN:1000-3673
Received: 06-01-2026 Revised: 15-02-2026 Accepted: 12-03-2026

3.5.2.1. Count Vectorizer

As depicted in Figure 13, the confusion matrix represents the performance of the model trained using the count vectorizer and Passive Aggressive Classifier on the raw dataset.

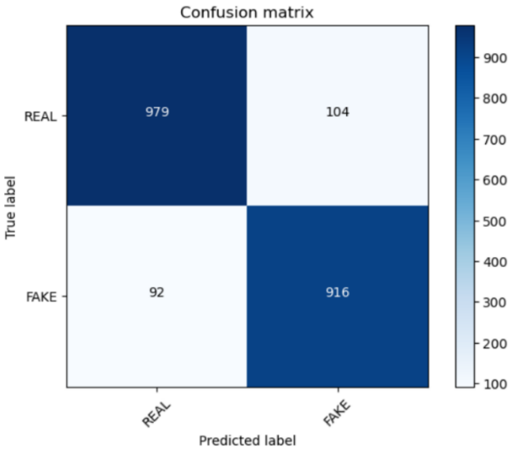


Figure 13. Count vectorizer and Passive Aggressive Classifier (Raw dataset)

As depicted in Figure 14, the confusion matrix represents the performance of the model trained using the count vectorizer and Passive Aggressive Classifier on the pre-processed dataset

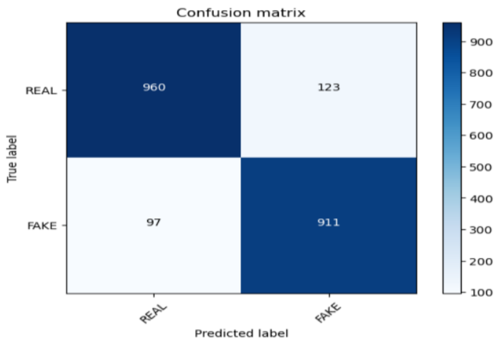


Figure 14. Count vectorizer and Passive Aggressive Classifier (Pre-processed dataset)

3.5.2.2. TF-IDF Vectorizer

As depicted in Figure 15, the confusion matrix represents the performance of the model trained using the TF-IDF vectorizer and Passive Aggressive Classifier on the raw dataset.

Power System Technology
ISSN:1000-3673
Received: 06-01-2026 Revised: 15-02-2026 Accepted: 12-03-2026

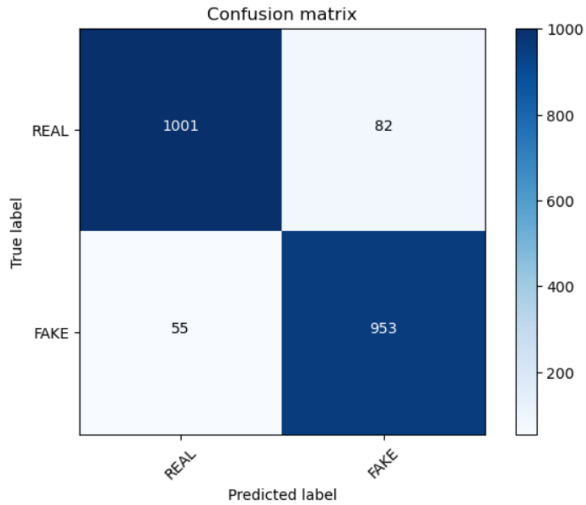


Figure 15. TF-IDF vectorizer and Passive Aggressive Classifier (Raw dataset)

As depicted in Figure 16, the confusion matrix represents the performance of the model trained using the TF-IDF vectorizer and Passive Aggressive Classifier on the pre-processed dataset.

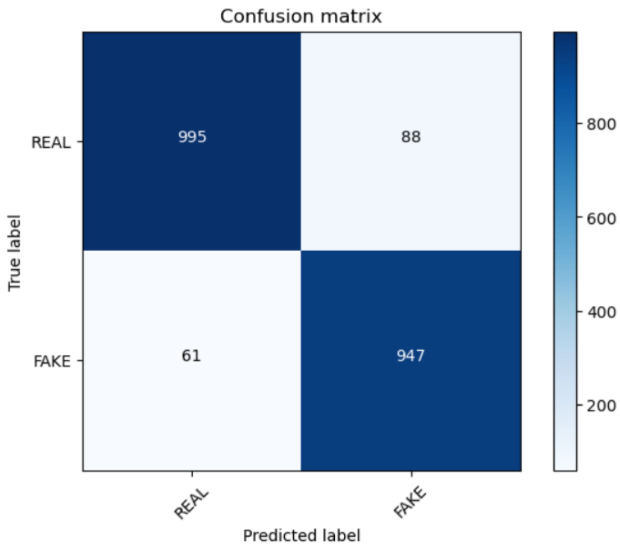


Figure 16. TF-IDF vectorizer and Passive Aggressive Classifier (Pre-processed dataset)

3.6 Model evaluation

The evaluation of the model is conducted using a test dataset. Various metrics, such as accuracy, precision, recall, and F1-Score, are computed to assess its performance.

3.6.1. Accuracy

Power System Technology
ISSN:1000-3673
Received: 06-01-2026 Revised: 15-02-2026 Accepted: 12-03-2026

The most widely used metric for precisely predicting the veracity of a statement is accuracy. The subsequent formula can be employed to ascertain the accuracy of a model.

Accuracy = (TP+FP)/(TP+FP+TN+FN)

3.6.2. Precision

Precision is a metric that gauges the quality of positive predictions made by a model. In our experiment, precision is calculated by dividing the number of true positives by the total number of positive predictions.

Precision = TP/(TP+FP)

3.6.3. Recall

Recall is defined as the total of correctly classified cases outside of the true class. In our experiment, we are especially addressing the number of articles among all accurately predicted articles that were correctly expected.

Recall = TP/(TP+FN)

Here, TP= True Positive, TN= True Negative, FP= False Positive, FN= False Negative

3.6.4. F-1 Score

In the context of our experiment, the F1 score provides a balanced measure that considers both the precision and recall of the model's predictions

F1-Score = 2 * (Precision * Recall) / (Precision + Recall)

**4. Results**

In this segment, we elaborate on the findings derived from the experiments conducted as part of the research study. We systematically compared different combinations of text vectorization techniques, specifically Count Vectorizer and TF-IDF Vectorizer, along with classifiers, including Multinomial Naïve Bayes Classifier and Passive Aggressive Classifier. This comparative analysis was performed on both processed and raw datasets, with evaluations based on key metrics such as accuracy, precision, recall, and F1-score.

4.1. Based on accuracy

This subsection focuses on the performance assessment of the evaluated combinations of text vectorization techniques and classifiers, with accuracy serving as the primary metric.

As shown in figure 17, on the raw dataset, the accuracy scores for models using Count Vectorizer with Multinomial Naïve Bayes (NB) classifier and Passive Aggressive classifier are 89.19% and 90.15%, respectively. Meanwhile, models employing TF-IDF with Multinomial NB classifier and Passive Aggressive classifier exhibit accuracies of 85.89% and

Power System Technology
ISSN:1000-3673
Received: 06-01-2026 Revised: 15-02-2026 Accepted: 12-03-2026

93.78%, respectively. The Passive Aggressive classifier with TF-IDF achieves the highest accuracy at 93.78%.

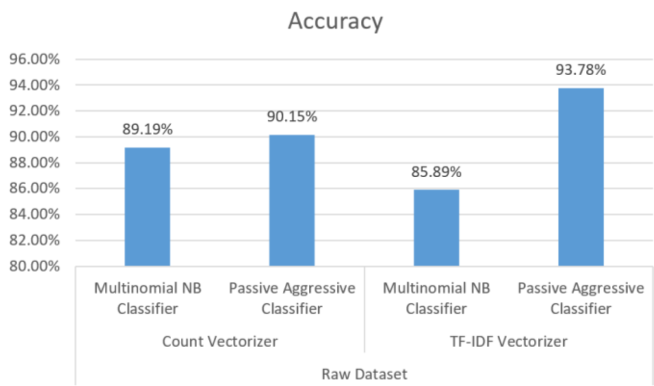


Figure 17. Accuracy vs Vectorization Techniques and Classifiers (On Raw Dataset)

As shown in figure 18, on the pre-processed dataset, the accuracy scores for models using Count Vectorizer with Multinomial Naïve Bayes (NB) classifier and Passive Aggressive classifier are 87.99% and 89.81%, respectively. Meanwhile, models employing TF-IDF with Multinomial NB classifier and Passive Aggressive classifier exhibit accuracies of 85.22% and 92.87%, respectively

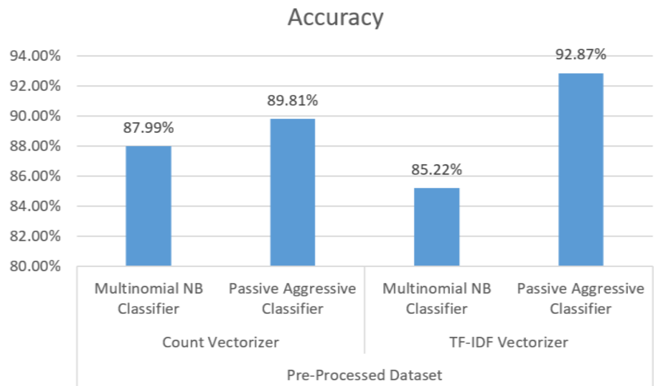


Figure 18. Accuracy vs Vectorization Techniques and Classifiers (On pre-processed Dataset)

4.2. Based on precision

This subsection focuses on the performance assessment of the evaluated combinations of text vectorization techniques and classifiers, with precision serving as the primary metric.

As shown in figure 19, on the raw dataset, using Count Vectorizer with Multinomial Naïve Bayes (NB) classifier and Passive Aggressive classifier, the precision scores are 87.42% and

Power System Technology
ISSN:1000-3673
Received: 06-01-2026 Revised: 15-02-2026 Accepted: 12-03-2026

91.02%, respectively. Meanwhile, employing TF-IDF with Multinomial NB classifier and Passive Aggressive classifier yields precision scores of 79.89% and 94.91%, respectively.

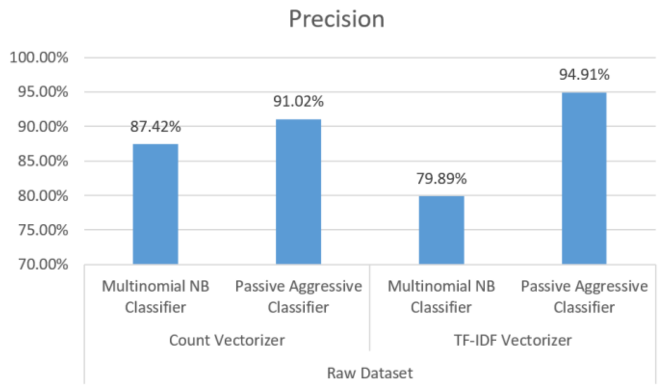


Figure 19. Precision vs Vectorization Techniques and Classifiers (On Raw Dataset)

As shown in figure 20, on the pre-processed dataset, using Count Vectorizer with Multinomial Naïve Bayes (NB) classifier and Passive Aggressive classifier, the precision scores are 86.81% and 91.12%, respectively. Meanwhile, employing TF-IDF with Multinomial NB classifier and Passive Aggressive classifier yields precision scores of 79.36% and 94.06%, respectively.

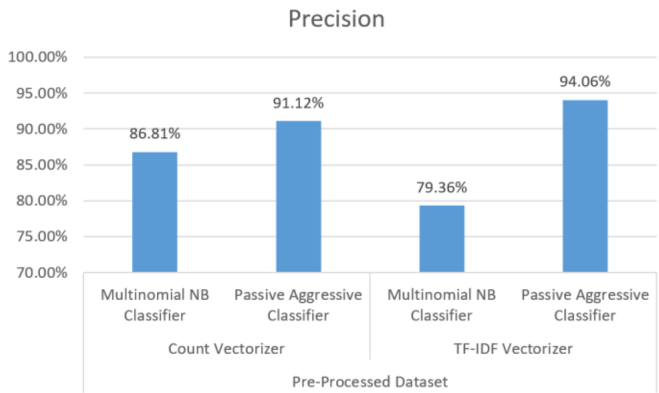


Figure 20. Precision vs Vectorization Techniques and Classifiers (On pre-processed Dataset)

4.3. Based on recall

This subsection focuses on the performance assessment of the evaluated combinations of text vectorization techniques and classifiers, with recall serving as the primary metric.

As shown in figure 21, on the raw dataset, using Count Vectorizer with Multinomial Naïve Bayes (NB) classifier and Passive Aggressive classifier, the recall scores are 92.43% and 89.84%, respectively. Meanwhile, employing TF-IDF with Multinomial NB classifier and Passive Aggressive classifier yields recall scores of 97.23% and 92.98%, respectively.

Power System Technology
ISSN:1000-3673
Received: 06-01-2026 Revised: 15-02-2026 Accepted: 12-03-2026

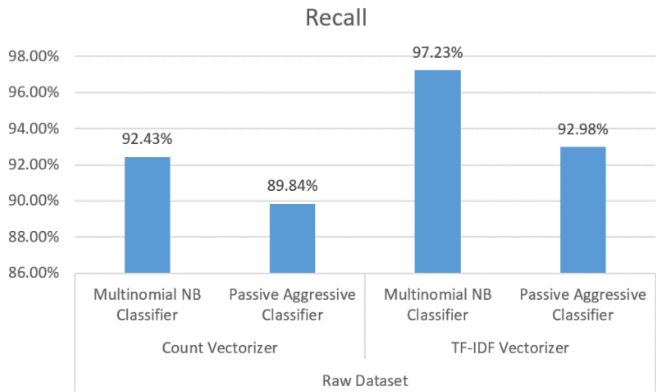


Figure 21. Recall vs Vectorization Techniques and Classifiers (On Raw Dataset)

As shown in figure 22, on the pre-processed dataset, using Count Vectorizer with Multinomial Naïve Bayes (NB) classifier and Passive Aggressive classifier, the recall scores are 90.58% and 89.04%, respectively. Meanwhile, employing TF-IDF with Multinomial NB classifier and Passive Aggressive classifier yields recall scores of 96.58% and 92.06%, respectively.

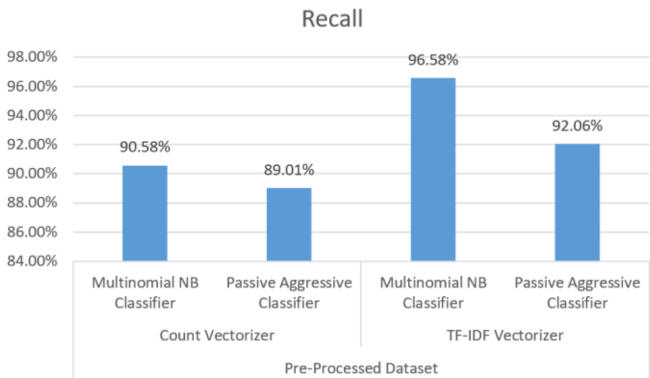


Figure 22. Recall vs Vectorization Techniques and Classifiers (On pre-processed Dataset)

4.4. Based on F-1 score

This subsection focuses on the performance assessment of the evaluated combinations of text vectorization techniques and classifiers, with f1-score serving as the primary metric.

As shown in figure 23, on the raw dataset, using Count Vectorizer with Multinomial Naïve Bayes (NB) classifier and Passive Aggressive classifier, the F1-scores are 89.86% and 90.43%, respectively. Meanwhile, employing TF-IDF with Multinomial NB classifier and Passive Aggressive classifier yields recall scores of 87.73% and 93.94%, respectively.

Power System Technology
ISSN:1000-3673
Received: 06-01-2026 Revised: 15-02-2026 Accepted: 12-03-2026

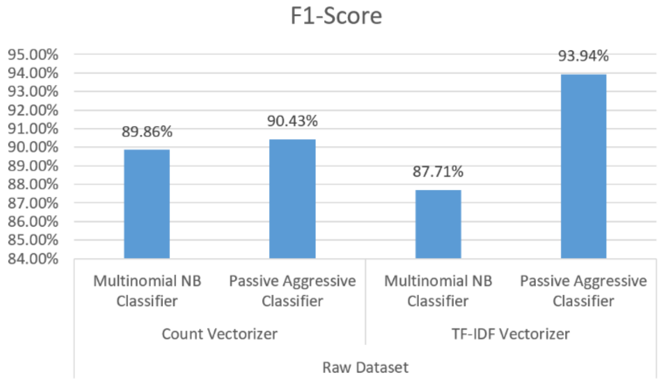


Figure 23. F1-Score vs Vectorization Techniques and Classifiers (On Raw Dataset)

As shown in figure 24, on the raw dataset, using Count Vectorizer with Multinomial Naïve Bayes (NB) classifier and Passive Aggressive classifier, the F1-scores are 86.66% and 90.05%, respectively. Meanwhile, employing TF-IDF with Multinomial NB classifier and Passive Aggressive classifier yields recall scores of 87.13% and 93.05%, respectively.

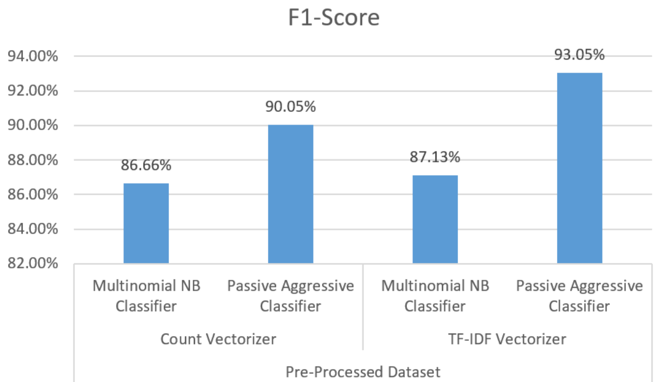


Figure 24. F1-Score vs Vectorization Techniques and Classifiers (On pre-processed Dataset)

4.5 Summarization

The following table is representing all the experiment results' accuracy, precision, recall, and f1-score mentioned below in figure 25.

| | Raw Dataset | | | | Pre-Processed Dataset | | | |
|---|---|---|---|---|---|---|---|---|
| | Count Vectorizer | | TF-IDF Vectorizer | | Count Vectorizer | | TF-IDF Vectorizer | |
| | Multinomial NB Classifier | Passive Aggressive Classifier | Multinomial NB Classifier | Passive Aggressive Classifier | Multinomial NB Classifier | Passive Aggressive Classifier | Multinomial NB Classifier | Passive Aggressive Classifier |
| Accuracy | 89.19% | 90.15% | 85.89% | 93.78% | 87.99% | 89.81% | 85.22% | 92.87% |
| Precision | 87.42% | 91.02% | 79.89% | 94.91% | 86.81% | 91.12% | 79.36% | 94.06% |
| Recall | 92.43% | 89.84% | 97.23% | 92.98% | 90.58% | 89.01% | 96.58% | 92.06% |
| F1-Score | 89.86% | 90.43% | 87.71% | 93.94% | 86.66% | 90.05% | 87.13% | 93.05% |

Figure 25. Summarization table

Power System Technology
ISSN:1000-3673
Received: 06-01-2026 Revised: 15-02-2026 Accepted: 12-03-2026

Upon careful examination of the presented table, it is evident that the combination of TF-IDF with the Passive Aggressive classifier achieves superior accuracy, precision, and f1-score when compared to other combinations of vectorization and classifiers. Consequently, both on raw and pre-processed datasets, this specific combination stands out as the most effective among the various text vectorization and classifier pairings explored in our study.

## 5. Conclusion

In conclusion, the comprehensive examination of text vectorization techniques and classifiers in the realm of fake news detection has yielded valuable insights. Through a systematic comparison of Count Vectorizer and TF-IDF Vectorizer, in conjunction with Multinomial Naïve Bayes and Passive Aggressive classifiers, on both raw and processed datasets, a nuanced understanding of their performance metrics has been achieved.

The findings underscore the significance of the TF-IDF vectorization method when paired with the Passive Aggressive classifier. Notably, this combination consistently outperforms other configurations, exhibiting superior accuracy scores, precision, recall, and F1-scores. On the raw dataset, the Passive Aggressive classifier with TF-IDF achieves the highest accuracy at 93.78%, marking it as a particularly robust choice for fake news detection.

This research emphasizes the importance of thoughtful consideration when selecting text vectorization techniques and classifiers for tackling the complex task of identifying fake news. The standout performance of TF-IDF with the Passive Aggressive classifier suggests that this pairing is not only effective on raw data but also maintains its efficacy on pre-processed datasets. As a result, researchers and practitioners in the field of fake news detection may find this specific combination to be a promising avenue for further exploration and application.

While the study does not claim an exhaustive evaluation of all possible combinations, the presented results provide valuable guidance for researchers and practitioners seeking optimized approaches in fake news detection. The ongoing evolution of misinformation highlights the need for robust methods, and the insights from this research contribute to the collective effort in developing more effective tools for combating the proliferation of fake news in our information landscape.

## 6. Future scope

The current research demonstrates promising results in detecting fake news using traditional text vectorization and classification algorithms. However, there remains significant potential for enhancement and expansion. Future work can explore the integration of advanced deep learning models such as LSTM, Bi-LSTM, and Transformer-based architectures to capture contextual and semantic relationships more effectively.

Power System Technology
ISSN:1000-3673
Received: 06-01-2026 Revised: 15-02-2026 Accepted: 12-03-2026

Additionally, incorporating multimodal data including images, videos, and metadata can enhance the robustness of fake news detection beyond text-based analysis. Expanding the dataset to include multilingual and cross-domain **sources** will improve the model's generalizability and adaptability across different regions and platforms.

Another promising direction is the development of real-time fake news detection systems capable of continuously monitoring and flagging unreliable content on social media. Further research could also investigate explainable AI (XAI) methods to make classification decisions more transparent and interpretable.

**References**

[1] Z Khanam, B N Alwasel, H Sirafi and M Rashid."Fake News Detection Using Machine Learning Approaches", IOP Conference Series Materials Science and Engineering, pp. 22–36, 2021. DOI:10.1088/1757- 899X/1099/1/012040

[2] A.Jain, H. Khatter, A. Shakya and A.K. Gupta. "A smart System for Fake News Detection Using Machine Learning", International Conference on Issues and Challenges in Intelligent Computing Techniques (ICICT), pp. 1-4, 2019. DOI: 10.1109/ICICT46931.2019.8977659.

[3] N. Fahad, K. O. Michael Goh, Md. I. Hossen, K. M. S. Shopnil and I. Jahan Mitu. "Stand up Against Bad Intended News: An Approach to Detect Fake News using Machine Learning" Emerging Science Journal, vol. 7, no. 4, pp. 1247-1259, 2023. DOI:10.28991/ESJ2023-07-04-015

[4] M. J. Awan, A. Yasin, H. Nobanee, A. A. Ali, Z. Shahzad and M. Nabeel."Fake News Data Exploration and Analytics", Electronics, vol. 10, no. 19, pp. 1-15, 2021. DOI:10.3390/electronics10192326

[5] R.K. Kaliyar, A. Goswami and P. Narang. " FakeBERT: Fake news detection in social media with a BERT-based deep learning approach" Multimedia Tools and Applications, Vol. 80. no 8, pp. 13439–13461, 2021.

[6] M. C Buzea, S. Trausan-Matu and T. Rebedea. "Automatic Fake News Detection for Romanian Online News" Information, Vol. 13, no. 3, pp. 1-13, 2022. DOI:10.3390/info13030151