EAI Endorsed Transactions
on AI and Robotics

Research Article EAI.EU

# A Comparison of the Performance of Six Machine Learning Algorithms for Fake News

Rafah H. Al-Furaiji[1,*] and Hasan Abdulkader[2]

[1]Electrical and Computer Engineering Department, Altinbas University, 34217 Istanbul, Turkiye
[2]Department of Computer Engineering, Altinbas University, 34217 Istanbul, Turkiye

## Abstract

**INTRODUCTION**: This research focuses on the increasing importance of social media websites as versatile platforms for entertainment, work, communication, commerce, and accessing global news. However, it emphasizes the need to use this power responsibly.
**OBJECTIVES**: The objective of the study is to evaluate the performance of artificial intelligence algorithms in detecting fake news.
**METHODS**: Through a comparison of six machine learning algorithms and the use of natural language processing techniques,
**RESULTS**: The study identifies four algorithms with a 99% accuracy rate in detecting fake news.
**CONCLUSION**: The results demonstrate the effectiveness of the proposed method in enhancing the performance of artificial intelligence algorithms in addressing the problem of fake news detection.

**Keywords**: Machine Learning, Natural Language Processing, Fake News, Scikit-Learn, Python

Received on 15 October 2023, accepted on 10 March 2024, published on 20 MM YYYY



doi: 10.4108/airo.4153

[1]Corresponding author. Email: 213720464@ogr.altinbas.edu.tr

## Introduction

The phenomenon of spreading fake news is one of the negative uses that technology and social media websites have been harnessed for. With these websites' ability to reach the farthest corners of the earth, it has become important to seriously address this problem and work on finding real solutions to reduce the great damage caused by this phenomenon.
In this research, we compared the performance of a group of machine learning models, in addition to using a group of natural language processing algorithms to process texts. We tried to shed light on some rarely used algorithms to verify their effectiveness in detecting whether the news is fake or real.

## 1. Related Work

In the research paper [1],[2], and [3], the researchers used a set of machine learning algorithms including Decision Tree and Logistic Regression, and some natural language processing methods. In both studies [1], and [3], the proposed models were trained on four manually classified datasets. The researchers in paper [1] focused on the importance of finding a comprehensive algorithm capable of dealing with all types of trained and untrained data with the same efficiency. Meanwhile, the researchers in the paper [**2**] suggested using tools such as Python scikit-Learn and NLP and a set of supervised machine learning algorithms to build a model capable of classifying news into true or false with the appropriate speed and efficiency for the current development. In paper [**3**], the researchers believe that many fake news stories contain many facts, and conversely, many true news stories contain a lot of false information. Therefore, the researchers evaluated the performance of four machine learning algorithms that work on detecting fake news with the help of natural language processing methods, and after training, the results were as shown in the table below:

Table 1: accuracy score paper [1]

| Model | DS1 | DS2 | DS3 | DS4 |
|---|---|---|---|---|
| Logistic Regression | 0.97 | 0.91 | 0.91 | 0.87 |
| Decision Tree | 0.98 | 0.94 | 0.94 | 0.9 |

Table 2: accuracy score paper [**3**]

| Model | K-FOLD-1 | K-FOLD-**2** | K-FOLD-**3** | K-FOLD-**4** |
|---|---|---|---|---|
| Decision Trees | 0.94 | 0.91 | 0.96 | 0.95 |

Paper [**4**] presents research on various machine learning algorithms, both individual classification algorithms and collaborative learning algorithms. The study also incorporates natural language processing techniques to identify the optimal algorithm for detecting fake news. Specifically, the paper explores the effectiveness of Logistic Regression, Decision Trees, Random Forest, SVC, AdaBoost, and XGBoost in this task.

The research paper [**5**] presents a solution for detecting fake news in Spanish by using traditional text feature extraction tools like TF-IDF and Stack, and a weak classifier-based ensemble learning approach. The proposed method extracts additional information about the text, such as information about publishers and topics published, and uses a variety of machine learning algorithms such as Logistic Regression, Passive Aggressive Classifier (PAC), SGD Classifier, and Ridge Classifier (RC). The test and validation datasets results were competitive, with an accuracy rate in IberLEF 2021 that was 8% higher than in MEX-A3T 2020. The researchers participated in the IberLEF 2021 campaign, which evaluates systems that process natural language in Spanish and other Iberian languages.

Table 3: accuracy score paper [**5**]

| Model | Merge All | Stop words | Accuracy |
|---|---|---|---|
| Logistic Regression (LR) | Yas | No | 0.859 |
| SGD Classier (SGDC) | Yas | No | 0.865 |
| Passive Aggressive Classier (PAC) | Yas | No | 0.870 |
| Ridge Classier(RC) | Yas | No | 0.862 |

In research papers [**6**] and [**7**], the focus was on machine learning algorithms such as Passive Aggressive classifier and Logistic regression that were applied in a set of algorithms to determine the best algorithm for detecting fake news. In the paper [6], four machine learning algorithms were used, including the Naïve Bayes algorithm, Support vector machine classification, Logistic regression, **Passive Aggressive classifier, and** some natural language processing techniques such as a bag of words and TD-IDF. The practical application results showed that the Passive Aggressive classifier gave the highest accuracy rate of 99.5%. In **the** paper [7], the researchers sought to determine whether the idea of using artificial intelligence to solve the problem of fake news is valid or not by creating a website that can help users verify fake news. Three machine learning algorithms were used, namely Logistic Regression, Naïve Bayes, and Random Forest, as well as some natural language processing techniques, and the results are shown in the table below:

Table 4: accuracy score paper [**6**]

| Model | Accuracy |
|---|---|
| Logistic Regression | 0.98 |
| Passive Aggressive classifier | 0.99 |

Table 5: accuracy score paper [**7**]

| Model | Accuracy |
|---|---|
| Logistic Regression | 0.65 |
| Passive Aggressive classifier | 0.92 |

In the paper [8] three models were utilized, including RoBERTa, a pre-trained language model from the BERT family, Bi-LSTM, a deep learning algorithm, and Passive Aggressive Classifier, a machine learning technique. It was discovered that RoBERTa delivered the best results with monolingual texts after choosing the right textual attributes and contrasting the results obtained by the three models. Bi-LSTM, however, performed better when dealing with texts in multiple languages.

Janicka et al. [**9**], a system design that may be applied to real-time news accuracy prediction is proposed. Data features are extracted from the data using natural language processing, and machine learning classifiers like Naive Bayes, Support Vector Machine (SVM), Random Forest (RF), Stochastic Gradient Descent (SGD), and Logistic Regression are trained using these features (LR). Many criteria are used to assess each classifier's performance. Then, the best classifier is implemented as a web application using the Flask API to predict the accuracy of news in real-time.

In the paper [**10**], the researchers focused on providing an answer to an important and fundamental question, which is

whether it is possible to build a model capable of handling all types of data that it has been trained on and those that it has not been trained on with the same efficiency. Based on this, the researchers built a model based on analyzing non-linguistic features such as writing style and psychological factors using four machine learning algorithms, including LinearSVC, SGD Classifier, Extra Trees Classifier, and XGBoost. This model was trained on four sets of data and the results showed that the performance of the model depends on the textual features and characteristics that it was trained on, and any interaction with other types of data reduces the efficiency of the model by 20%. Therefore, training the model on data sets from various domains is important.

In the paper [11], machine learning and deep learning algorithms were employed to ascertain the accuracy of news by determining its veracity. Two methodologies, both centered around language, were utilized in this study. The first approach involved concatenating the news title and major news material for the experiment. The second approach focused solely on the news content, disregarding the title.

In [12] [13], the researchers used Logistic Regression and Naive Bayes machine learning algorithms to detect fake political news. In the paper [12], the algorithms were applied in two stages: the first stage involved detecting fake news, while the second stage was verifying fake news detection. In paper [13], a framework for detecting fake news has been proposed based on feature extraction and selection techniques such as inverse document frequency and bag of words, as well as a set of classifiers including voting classifiers. The extracted features were reduced using various machine learning algorithms and a variance analysis algorithm.

In **the** paper [14], the researchers worked on discovering fake news on the Twitter platform. However, tweets on Twitter are short texts that are not subject to any syntactic rules. Therefore, the researchers proposed using tweet features such as the number of retweets, likes, and tweet length. They trained a set of algorithms including SVM, decision tree, and neural network.

Adedoyin et al. [15], the performance of seven algorithms was evaluated, including five machine learning algorithms and two deep learning algorithms, in the field of detecting fake news on social media platforms. Using a set of measurement tools, the best model was determined, and a web application was built for this purpose.

## 2. Methodology

In this research paper, we will evaluate and compare the performance of six machine learning algorithms to determine the best algorithm for detecting fake news. These algorithms are Extra Tree Classifier, Logistic Regression, SGD Classifier, Passive Aggressive Classifier, Ridge Classifier, and Decision Tree Classifier, with the assistance of some natural language processing tools. The selected algorithms are provided by the scikit-learn (sklearn) library written in Python, and they are among more than 20 other algorithms available in this library for text processing and classification. In our research, some of the selected algorithms are less well-known, as there are not many studies that discuss these algorithms, and we did not find any previous study comparing the six algorithms we chose for this research. Therefore, we took the initiative to compare these six algorithms.

In this section, we have four stages to reach the desired results. In the first stage, we analyze the data. In the second stage, we process the data and convert it into vectors. In the third stage, we train the algorithm on the training data. Finally, in the fourth stage, we evaluate the performance of the algorithm on the test data.

### A. Data Analysis

Using visual and analytical analysis, it's possible to identify and verify fake news effectively. Therefore, we analyze the data set by displaying the most important words using a word cloud tool, in addition to displaying the top 20 most frequently repeated words in the data set using Count Vectorizer.

**A.1 word cloud:** A visual representation of the frequency of words in a text where the size of the word denotes the frequency.

**A.2 Top 20 Words:** By counting the top 20 most frequent words, Count Vectorizer can offer insightful information about the textual data that has been studied. This can assist in understanding the words that appear in the text the most frequently as well as in locating relevant patterns and key terms**.**

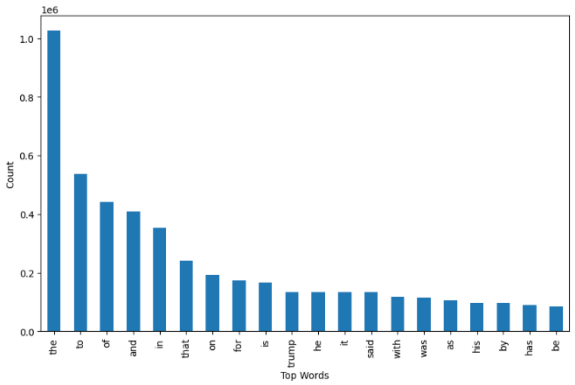


**Figure 1.** Bar Chart of Top Words Frequency.

### B. Data pre-processing

Text written in natural language is often chaotic and noisy because it contains so much unimportant information. Text preparation is required to get it ready for additional investigation and education. The text must be transformed into an organized and dependable format to feed the content into a

model. In this research, we utilized three natural language processing tools, which are:

- remove irrelevant spaces,
- remove punctuations mark,
- remove the stopwords,
- split data into training and test.

**B.1 Remove irrelevant spaces:** To ensure data cleanliness, it is crucial to eliminate any additional spaces and eradicate newlines, tabs, and any form of white space from the given dataset.

**B.2 Remove punctuation marks:** When working with machine learning and data processing tasks, it's common to remove punctuation as a preliminary step.

**B.3 Remove the stop words:** We try to stay away from any phrases that use up extra resources so that our database can be stored and processed as efficiently as possible. This can be accomplished by keeping a list of stop words, or words with limited or no meaning in a particular language. For instance, NLTK offers a set of stop words in 16 different languages that we can use.

**B.4 Split data into train and test:** As a pre-processing step in machine learning, dividing the modeling dataset into training and testing samples is essential. The performance of the model can be evaluated by generating multiple training and testing samples.

## C. Converting text into Vectors

Text vectorization converts textual data into a numerical format, enabling computers to understand and handle the input. In this paper, we will use TF-IDF to convert text into a numerical format and to understand what TF-IDF it is first necessary to define term frequency (TF) and inverse document frequency (IDF) separately. Let's first get acquainted with the term frequency (TF). It is a term that, said simply, denotes the weightiness of a word in a document, The term inverse document frequency (IDF) represents the weight of a word across all documents.
The following steps are used in the TF-IDF technique to convert text into numerical representations:

- Calculation of Term Frequency (TF): Based on the frequency of each word in the text, TF generates a numerical value. It shows how frequently a word appears in the text.

  **TF (t)= (number of times word t appears in the text) / (total number of words in the text)**

  the formula for calculating TF

- Calculating inverse document frequency (IDF) measures a word's importance across the full corpus of documents. It displays the frequency with which the word appears across all of the publications.

  **IDF(t) = log((total documents in the collection) / (documents with the phrase t in them))**

  the formula for computing IDF

- Calculating Term Frequency-Inverse Document Frequency (TF-IDF): By dividing a word's TF value by its IDF value, the TF-IDF is calculated. For the full collection of papers, it represents the normalized phrase frequency. TF-IDF is calculated using the following formula:

  **TF-IDF(t) = TF(t) * IDF(t)**

  the formula for computing TF_IDF

These procedures transform the text into numerical vectors that correspond to the document's word representations. Higher TF-IDF values imply that the documents contain unique and important words.

## D. Train the Models

In this research paper, we have six supervised learning algorithms that will be applied to two manually classified datasets to evaluate the performance of these algorithms and determine the best algorithm for solving the problem of fake news among the six algorithms.

**D.1 Logistic Regression (LR):** For categorization and predictive analytics, this form of statistical model, often called a logistic regression model, is frequently employed. Based on a specific dataset of independent variables, logistic regression calculates the probability of an event occurring, such as voting or not voting. The dependent variable is limited to the range of 0 and 1, as the result is a probability.

**D.2 SGD Classier (SGDC):** For creating linear classifiers and calculating gradients under convex loss functions, such as logistic regression and linear classifiers, stochastic gradient descent (SGD) is seen to be a straightforward and effective method. The SGD methodology is essentially a method for training models; it is not connected to any particular family of machine learning models. It has been effectively used to solve a wide variety of **machine-learning** issues, including text categorization and natural language understanding.

**D.3 Passive Aggressive Classier (PAC):** The online learning algorithms encompass various approaches, one of which is the Negative Aggressive Learning Algorithm. This algorithm exhibits a unique behavior where it responds negatively when it correctly classifies data and reacts strongly when it makes incorrect classifications. Specifically designed for online machine learning, the negative aggressive algorithm learns gradually by processing consecutive examples. This learning process can be performed individually or by grouping small samples into mini batches. The primary objective of the negative aggressive algorithm is to become proficient at identifying accurate classifications and responding negatively to them. Upon receiving a new sample, the algorithm assesses whether it belongs to the correct classification and reacts aggressively if an incorrect judgment is made. To enhance its capabilities, the negative aggressive algorithm undergoes continuous training. This training aims to improve its ability to identify accurate classifications and enhance its performance in handling inaccurate classifications.

**D.4 Ridge Classier (RC):** This is a method used in machine learning to examine linear discriminant models. To avoid overfitting, this type of regularization penalizes model coefficients. When a model is overly complicated and captures noise in the data rather than the underlying signal, it is known as overfitting, a prevalent problem in machine learning. On new data, this may result in subpar generalization performance. Ridge classification tackles this issue by including a penalty term that deters complexity in the cost function. As a result, the model's ability to generalize to new data is improved. The goal values are first transformed into the values "-1," and "1," and the problem is then handled as a regression task (multi-output regression in the case of many classes).

**D.5 Decision Tree Classifier (DT):** Suitable for classification and regression tasks, decision trees (DTs) are a non-parametric supervised learning method. Their main goal is to build a model that can foretell the value of the target variable. This is accomplished by taking the simple decision rules from the data's features and applying them. A conceptual representation of a tree can be seen as an approximation made up of segments with piecewise constant length.

**D.6 Extra Tree Classifier (ET):** Extra trees, also known as extremely randomized trees, are an ensemble supervised machine learning technique that utilizes decision trees. It is commonly employed in the Train Using AutoML tool. This approach bears a resemblance to random forests but boasts potential speed advantages.

## 3. Results and Discussion

In this study, the quality of machine learning algorithms' performance in fake news detection was evaluated using performance indicators such as accuracy, recall, F1-score, and precision. These measurements were based on the analysis of the confusion matrix. The experimental results involved the application of six machine-learning algorithms to two manually trained datasets. The findings indicated that the performance of these algorithms in the field of fake news detection has reached advanced stages. Among the algorithms tested, the Extra Tree algorithm achieved a minimum accuracy rate of 86%. However, the remaining algorithms surpassed 99% accuracy, indicating their superior performance in this task.

Confusion Matrix: **An effective tool for visualizing the many outcomes of predictions and results in a classification task is the confusion matrix. It displays a table with all of a classifier's predicted and actual values. There are four primary sections in the Confusion matrix:**

- True Positive (TP): Indicates how many true positive samples the model properly identified.
- False Positive (FP): Indicates the number of samples that the model misclassified as positive and so counted as false positives.
- True Negative (TN): Indicates how many true negative samples the model properly identified.
- False Negative (FN): Indicates the number of samples that the model misclassified as negative and so counted as false negatives.

In this research, we find that the color shades in the cells of the Confusion matrix are very close. In our study, the matrix consists of two-color shades: pink and black. The pink color represents True Positive (TP) and True Negative (TN), while the black color represents False Positive (FP) and False Negative (FN). By following the obtained chart, we can see that the models perform very well, and this can be observed when examining the results using other measurement tools that rely primarily on matrix outcomes.

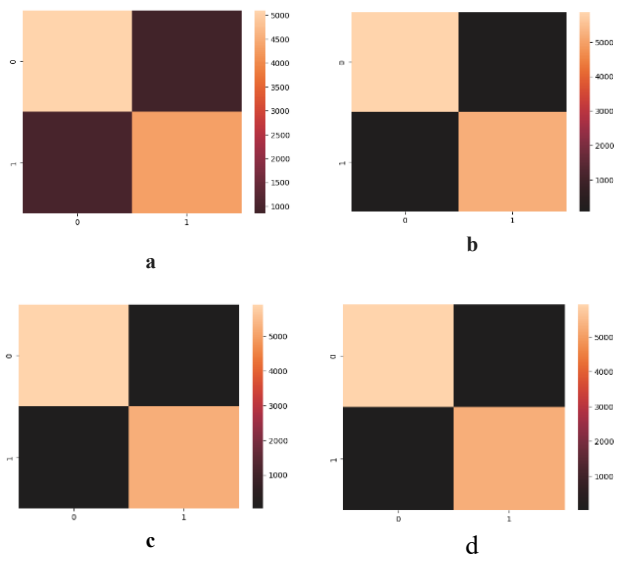


**Figure 2.** The Confusion Matrix for Dataset-1: a. ET, b. LR, c. RC, d. DTs

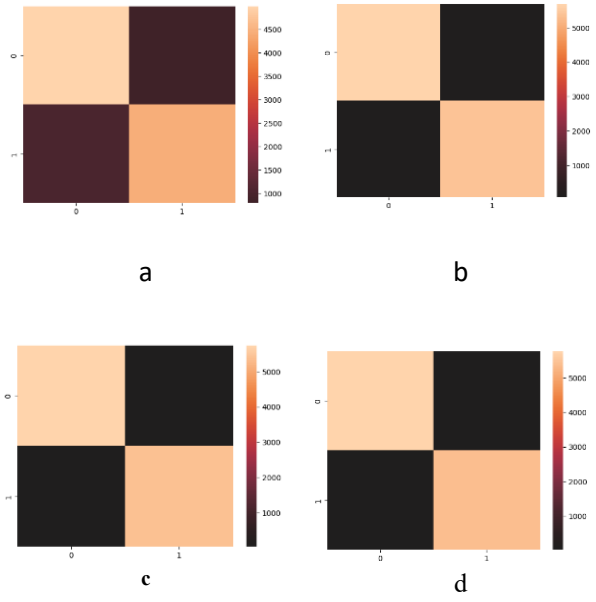


**Figure 3.** The Confusion Matrix for Dataset-2: a. ET, b. LR, c. RC, d. DTs

A. Accuracy: The percentage of accurate predictions made by the model is calculated using the accuracy metric in classification tasks. Machine learning defines accuracy as an evaluation metric that counts how many of the model's predictions were accurate and divides that number by the overall number of predictions. It is derived by dividing the total number of predictions by the number of accurate predictions.

B. Precision: The proportion of accurate positive forecasts to all positive predictions is known as the precision ratio. It shows how many predictions successfully identified the target variable as being positive. If the precision is less than 0.5, there are likely to be many more false positive predictions than true ones. In other words, accuracy refers to the capacity to correctly predict outcomes that are positive among all predictions that are made and categorized as positive. There are a lot of false positive predictions if the precision is less than 0.5.

C. Recall: The proportion of correct positive predictions to the total number of times the target variable was positive is known as recall. A high percentage of False Negatives is indicated by recall values that are less than 0.5.

D. F1 Score: By calculating the harmonic mean of the data, the F1 score is a statistic that combines the Precision and Recall values. In other words, it accounts for both cases of false positive (FP) and false negative (FN) results.

Table 6: accuracy score DS1

| Model | Accuracy | Precision | Recall | F1 Score |
|---|---|---|---|---|
| ET | 0.8435 | 0.8500 | 0.8162 | 0.8328 |
| LR | 0.9865 | 0.9833 | 0.9886 | 0.9859 |
| SGD | 0.9917 | 0.9912 | 0.9914 | 0.9913 |
| PAC | 0.9957 | 0.9962 | 0.9947 | 0.9955 |
| RC | 0.9948 | 0.9949 | 0.9942 | 0.9945 |
| DTs | 0.9943 | 0.9949 | 0.9930 | 0.9940 |

Table 7: accuracy score DS2

| Model | Accuracy | Precision | Recall | F1 Score |
|---|---|---|---|---|
| ET | 0.8626 | 0.8636 | 0.8407 | 0.8520 |
| LR | 0.9865 | 0.9819 | 0.9895 | 0.9857 |
| SGD | 0.9903 | 0.9870 | 0.9926 | 0.9898 |
| PAC | 0.9949 | 0.9933 | 0.9958 | 0.9946 |
| RC | 0.9940 | 0.9913 | 0.9960 | 0.9936 |
| DTs | 0.9959 | 0.9958 | 0.9956 | 0.9957 |

To demonstrate the magnitude of the developments in the performance of these algorithms, we did a simple comparison between the findings we obtained and the results of earlier studies.

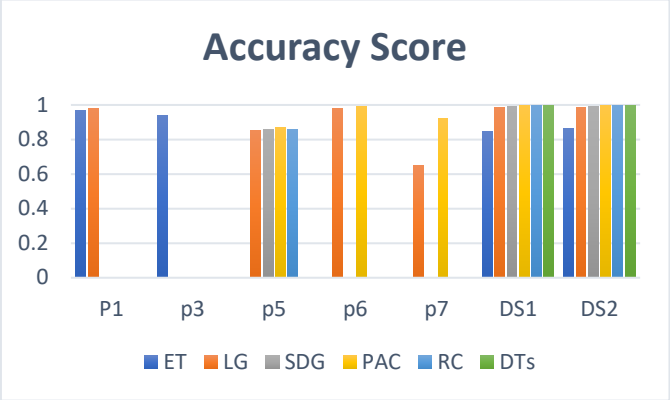


**Figure 4.** Accuracy Score

Figure 4 compares our obtained results against state of art. P1 to p7 are respectively papers [1] – [7].

## Conclusion

In this research paper, a system for the detection of fake news is implemented using machine learning and natural language processing. Six different machine learning were implemented, and their results are compared using several metrics. Metrics based on 2 different datasets showed ML achieve accuracy rates ranging from 86% to 99%. Four of these algorithms achieved very high accuracy rates, reaching up to 99%, with very small fractional differences between them. These algorithms are SGD, PAC, RC, and DTs. Following them is the LR algorithm, which achieved an accuracy rate of 98%. Finally, the ET al algorithm yielded the lowest accuracy rate at 86%, which is still not considered a poor percentage. This demonstrates the significant advancements in the use of artificial intelligence and natural language processing techniques in detecting fake news.

## Reference


[1] Iftikhar Ahmad, Muhammad Yousaf, Suhail Yousaf , Muhammad Ovais Ahmad, "Fake News Detection Using Machine Learning Ensemble Methods," Hindawi, p. 11, 2020. https://doi.org/10.1155/2020/8885861

[2] 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), 012040. https://doi.org/10.1088/1757-899x/1099/1/012040

[3] Albahr, A., & Albahar, M. (2020). An Empirical Comparison of Fake News Detection using Different Machine Learning Algorithms. International Journal of Advanced Computer Science and Applications : IJACSA, 11(9). https://doi.org/10.14569/ijacsa.2020.0110917

[4] Singh, G., & Selva, K. (2023). Detection of fake news using NLP and various single and ensemble learning classifiers. https://doi.org/10.36227/techrxiv.21856671.v2

[5] HAHA at FakeDeS 2021: A Fake News Detection Method Based on TF-IDF and Ensemble Machine Learning. (n.d.).

[6] Thirupathi, L., Gangula, R., Ravikanti, S., Sowmya, J., & Shruthi, S. K. (2021). False news recognition using machine learning. Journal of Physics. Conference Series, 2089(1), 012049. https://doi.org/10.1088/1742-6596/2089/1/012049

[7] Sharma, , U., Saran, , S. and Patil, S.M. (2021) 'Fake News Detection using Machine Learning Algorithms', International Journal of Engineering Research & Technology (IJERT), pp. 509–518.

[8] Arif, M. et al. (2022) 'CIC at CheckThat! 2022: Multi-class and Cross-lingual Fake News Detection', CLEF 2022: Conference and Labs of the Evaluation Forum [Preprint]. Available at: https://ipn.elsevierpure.com/en/publications/cic-at-checkthat-2022-multi-class-and-cross-lingual-fake-news-det (Accessed: 2022).

[9] Janicka, M., Pszona, M., & Wawer, A. (2019). Cross-domain failures of fake news detection. Computación y Sistemas, 23(3). https://doi.org/10.13053/cys-23-3-3281

[10] Arif, M. et al. (2022) 'CIC at CheckThat! 2022: Multi-class and Cross-lingual Fake News Detection', CLEF 2022: Conference and Labs of the Evaluation Forum [Preprint]. Available at: https://ipn.elsevierpure.com/en/publications/cic-at-checkthat-2022-multi-class-and-cross-lingual-fake-news-det (Accessed: 2022).

[11] Team Sigmoid at CheckThat!2021 Task 3a: Multiclass fake news detection with Machine Learning. (n.d.).

[12] Sudhakar, M., & Kaliyamurthie, K. P. (2022). Effective prediction of fake news using two machine learning algorithms. Measurement. Sensors, 24(100495), 100495. https://doi.org/10.1016/j.measen.2022.100495

[13] Gaikwad, T., Rajale, B., Bhosale, P., Vedpathak, S., & Adagale, M. S. S. (2022). Detection of fake news using Machine Learning Algorithms. International Journal for Research in Applied Science and Engineering Technology, 10(12), 1506–1510. https://doi.org/10.22214/ijraset.2022.48172

[14] Diaa SALAMA ABD ELMINAAM, Yomna M.I.HASSAN, Radwa MOSTAFA, Abd Elrahman TOLBA ,Mariam KHALED, and John GERGES"Fake News Detector Using Machine Learning Algorithms" Proceedings of the 36th International Business Information Management Association (IBIMA), ISBN: 978-0-9998551-5-7, 4-5 November 2020,Granada, Spain.

[15] Adedoyin, F., & Mariyappan, B. (2022b). Fake news detection using machine learning algorithms and recurrent neural networks. In SAGE Advance. https://doi.org/10.31124/advance.20751379.v1