FAKE NEWS DETECTION USING DECISION TREE AND ADABOOST

Anuradha K[1], Senthil Kumar P[2], Naveen Prasath E[3], Vignesh M[4], Sneha S[5]

Abstract

The prevalence of fake news in social media and online platforms has become a growing concern for society. To address this issue, researchers have explored various techniques for fake news detection. The proposed method first preprocesses the news articles using Natural Language Processing (NLP) techniques, including tokenization, stop-word removal, and stemming, to extract relevant features from the text data. These features are then used as input to the decision tree algorithm to classify news articles as real or fake. In this study, we propose a method for fake news detection using decision trees and Ada boost. Decision trees are used to extract relevant features from the text data, and Ada boost is employed to enhance the performance of the decision trees. The proposed method is evaluated on a dataset of news articles, and the results demonstrate its effectiveness in detecting fake news. The proposed method can be used as a tool for identifying fake news and could be beneficial in addressing the issue of misinformation in online platforms. The experimental results show that the proposed method using decision trees achieves high accuracy in fake news detection, which demonstrates the effectiveness of decision trees in feature extraction. Furthermore, Ada boost is shown to improve the performance of the decision tree algorithm by adjusting the weights of misclassified samples, which further improves the overall classification performance. This paper evaluated the proposed method on a dataset of news articles and achieves an accuracy of 99% using decision trees and 82% using Ada boost. This paper demonstrates and analyzes the detection of fake news with popular machine learning algorithms.

Keywords: Machine learning,Decision tree, Adaboost, Fake news detection, unverified news, news classification.

[1]Director, Department of Master of Computer Applications, Karpagam College of Engineering, Coimbatore, India.
[2,3,4,5]Final MCA, Department of Master of Computer Applications, Karpagam College of Engineering, Coimbatore, India.

Email: [1]k_anur@yahoo.com, [2]senthilkumar2712001@gmail.com.

1. Introduction

In the 20th century, the Internet has become the most powerful tool for communication. It facilitates the efficient and effective transfer of media from one location to another. With the development of Internet technology, social networks such as Facebook, WhatsApp, Twitter and Instagram have become a vital platform for information exchange. Lie gets traveled around us quicker, and more extensively than reality in all spheres of information, and the effects were more dangerous and horrifying. As fast the technology is moving, on the same pace preventive measures are required to deal with such activities. Broad communications assume a gigantic job in impacting the general public and as it is normal, a few people attempt to exploit it. There are numerous sites that give false data. The main objective is to detect the fake news, which is a text classification problem. It is needed to build a model that can differentiate between "Real" and "Fake" news. This leads to consequences in social networking site, microblogging and instant messaging applications where these fake news gets a major boost and gets viral among people, around the world. In the age of technology, a tremendous amount of data is being generated online every day. However, an unprecedented amount of the data flooded on the Internet is fake news, which is generated to attract the audience, to influence beliefs and decisions of people to increase the revenue generated by clicking, and to affect major events such as political elections. Readers are misguided by deliberately spreading false information. The proliferation of fake news and misinformation on social media and online platforms has become a major concern in recent years. Fake news has been shown to have a significant impact on public opinion and can lead to negative consequences, such as social unrest and political instability. Therefore, the detection of fake news has become an essential task for media, government agencies, and the general public. Fake news is a growing problem in today's digital age, with the widespread use of social media and other online platforms providing easy access to information that can be easily manipulated or fabricated. The spread of fake news can have serious consequences, including social and political polarization, erosion of trust in institutions, and even physical harm in some cases. Therefore, the development of effective methods for detecting fake news has become a critical need. One approach to fake news detection is through the use of machine learning algorithms, which can be trained on large datasets of labeled news articles to automatically identify patterns and features that distinguish between real and fake news. There are several popular machine learning algorithms that

can be used for fake news detection, including decision tree, support vector machines, random forest, and neural networks.

Machine learning techniques have been extensively used to detect fake news. Decision tree and Adaboost are two popular machine learning algorithms that have been employed in this area. Decision trees are effective in feature extraction and have been used in many natural language processing (NLP) applications. Adaboost is an ensemble learning algorithm that combines multiple weak classifiers to create a strong classifier and has been shown to improve the performance of decision trees. Several studies have employed decision tree and Adaboost algorithms for fake news detection. Deep learning algorithms have several advantages over traditional machine learning algorithms in the context of fake news detection. They are capable of automatically learning features from raw data, which can be especially useful in cases where it is difficult to define a set of features a priori. Additionally, deep learning algorithms can model complex relationships between inputs and outputs, which can be valuable in the context of detecting subtle patterns and distinguishing between real and fake news. Several deep learning architectures have been used for fake news detection, including convolutional neural networks (CNNs), recurrent neural networks (RNNs), and transformers. CNNs are commonly used for image recognition tasks, but they can also be applied to natural language processing tasks, such as fake news detection. RNNs are a type of neural network that are well-suited for sequence modeling, making them useful for tasks that involve analyzing sequences of words, such as text classification. Transformers are a newer type of neural network architecture that have shown impressive results in natural language processing tasks, including fake news detection.The use of deep learning algorithms for fake news detection has shown promising results in recent studies, with many achieving high levels of accuracy. However, challenges remain, such as the need for large labeled datasets and the potential for deep learning models to overfit to training data. Despite these challenges, the use of deep learning algorithms offers a powerful and scalable approach to detecting fake news, with the potential to significantly mitigate the spread of misinformation in online platforms.

In one study, authors proposed a fake news detection method based on a decision tree and achieved an accuracy of 98.34%.[1] Decision tree algorithms was used to detect fake news and achieved good accuracy. [2]. Several study proposes a method for fake news detection using decision tree and Adaboost algorithms with NLP

techniques for feature extraction. The proposed method is an effective way to identify fake news and has the potential to be used as a tool for addressing the issue of misinformation in online platforms. The combination of NLP techniques and machine learning algorithms has shown promising results in identifying fake news. Studies have demonstrated that the use of decision tree and Adaboost algorithms in fake news detection can improve the accuracy of classification. Additionally, some studies have proposed hybrid approaches combining multiple machine learning algorithms for fake news detection, achieving high accuracy levels [3, 4]. The method using decision tree and Ada boost algorithms provides an effective way to detect fake news, and the use of NLP techniques improves the accuracy of feature extraction. The combination of these methods has shown promising results in identifying fake news and could be beneficial in addressing the issue of misinformation on online platforms.

Literature Survey
A strategy for identifying fake news on social media has been proposed by a study, and it entails examining features that are network-, user-, and content-based. To determine if news stories are authentic or not, the authors use a variety of machine learning methods, such as decision trees, random forests, and support vector machines. They illustrate the potency of their framework for identifying false news by evaluating it on a sizable dataset of news stories from Twitter [5]. For the purpose of identifying fake news, a research suggests a neural network architecture termed a Long Short-Term Memory (LSTM) network. The authors show that their method is effective in reaching high accuracy in their experiments by using a dataset of tagged news items. Additionally, they illustrate the effectiveness of their strategy in identifying bogus news by contrasting it with other machine learning algorithms [6]. Another study suggests a hybrid strategy to identify fake news that blends deep learning algorithms, such as convolutional neural networks (CNNs) and recurrent neural networks (RNNs), with NLP methods, such as sentiment analysis and word embedding. The authors show that their method is effective in reaching high accuracy in their trials by using a sizable dataset of news items. Additionally, they illustrate the effectiveness of their strategy in identifying bogus news by contrasting it with other machine learning algorithms [7].

A technique for detecting fake news called ensemble learning that integrates different machine learning algorithms like decision trees, random forests, and AdaBoost. The authors show that their method is effective in reaching high accuracy in their experiments by using a dataset of tagged

news items. Additionally, they compare their method to other machine learning algorithms and show how effective their method is in spotting bogus news[8]. A multi-modal deep learning system that incorporates text, image, and user data for fake news identification. The authors show that their method is effective in reaching high accuracy in their experiments by using a dataset of news items. Additionally, they illustrate the effectiveness of their strategy in identifying bogus news by contrasting it with other machine learning algorithms. [9]. An article offers a thorough analysis of deep learning methods for spotting fake news, including CNNs, RNNs, and transformers. The authors analyse current research on false news identification and evaluate the effectiveness of several deep learning methods. They also go over the difficulties and restrictions of applying deep learning to the identification of fake news, and they offer suggestions for future research. [10]. With an emphasis on the methodologies, strategies, and datasets employed in these investigations, this paper offers a summary of recent research on automated false news identification. The authors offer insights into the efficiency of various approaches, such as linguistic, social, and network-based methodologies, in identifying fake news by analysing the advantages and disadvantages of each. Additionally, they evaluate the drawbacks and shortcomings of current strategies and make suggestions for future research possibilities[11]. The suggests combining network analysis, sentiment analysis, and machine learning approaches in conjunction with data mining to identify bogus news on social media. The authors demonstrate the effectiveness of their approach in achieving high accuracy in false news detection by evaluating it on a sizable dataset of tweets. In addition, they evaluate the characteristics and algorithms employed in their methodology and recommend future research trajectories. The feature extraction method and the classifier Support Vector Machine (SVM). We also suggest a dataset of real and fraudulent news for the suggested system's training. Results obtained demonstrate the system's effectiveness[12]. The authors examine the difficulties and restrictions of current techniques and make recommendations for future study areas. They also go over the advantages and disadvantages of various machine learning algorithms, such as decision trees, random forests, and deep learning, and offer information on how well they can identify bogus news[13]. survey of current works on the identification of fake news on Twitter, with an emphasis on the datasets, features, and algorithms employed in these investigations. The authors assess the advantages and disadvantages of various machine learning methods and offer information on how well they can identify fake news[14]. Survey of recent

studies on fake news detection, with a focus on the fundamental principles and challenges of detecting fake news. The authors analyze the features and algorithms used in different studies and provide insights into their effectiveness in detecting fake news. They also discuss the challenges and limitations of existing approaches and suggest future research directions[15].

Deep learning approach for fake news detection that combines ensemble learning with adversarial training. The authors evaluate their approach on a large dataset of news articles and demonstrate its effectiveness in achieving high accuracy in detecting fake news[16]. a deep learning method that combines CNNs with attention mechanisms for the identification of bogus news. The authors show that their method is effective in reaching high accuracy in their experiments by using a dataset of news items. Additionally, they illustrate the effectiveness of their method for spotting fake news by comparing it to other machine learning algorithms[17]. A machine learning-based approach for detecting fake news. As a feature extraction strategy, employed term frequency-inverse document frequency (TF-IDF) of a sample of words and n-grams, and Support Vector Machine (SVM) as a classifier. We also suggest a dataset of real and fraudulent news for the suggested system's training. Results obtained demonstrate the system's effectiveness. In this work, we provide a machine learning-based approach for detecting fake news. As a feature extraction strategy, we employed term frequency-inverse document frequency (TF-IDF) of a sample of words and n-grams, and Support Vector Machine (SVM) as a classifier. We also suggest a dataset of real and fraudulent news for the suggested system's training. The outcomes demonstrate the effectiveness [18]. The dissemination of false information is widespread, the credibility of social media networks is also at jeopardy. So, it has become a research problem to automatically categorize material as accurate or false based on its source, substance, and publisher. Despite its limits, machine learning has been crucial in the classification of data. This research examines various machine learning techniques for the identification of produced and false news. It also reviews the limitations of these methods and improvisational deep learning implementation[19].

To recognize bogus news Yet, the amount of data on the internet and in social media is growing rapidly. Because it is difficult and time-consuming to sort through all of this data to determine if news is true or false, classification algorithms are used to sort this enormous amount of data. Here, suggested a classification-based fake news detection system using techniques including deep neural networks (DNN), support vector machines, naive bayes, logistic regression, and naive bayes For the purpose of identifying fake news, we compare every machine learning method[20]. It provides a thorough summary of the research that has been done so far on false news. It describes the detrimental effects of internet fake news as well as cutting-edge detecting techniques. Numerous of them rely on spotting user characteristics, content, and context that point to disinformation. It also look at datasets that have already been used to categorize false information. Lastly, we suggest prospective study avenues for the examination of online fake news[21]. In this article, provides a technique for identifying "fake news" and ways to use it on Facebook, one of the most well-known social networking websites. The Naive Bayes classification algorithm is used in this technique to forecast whether a Facebook post will be categorized as real or fake. Using a number of the approaches suggested in the study may improve the outcomes. According to the results, machine learning techniques can be used to solve the false news identification problem [22].

The paper offers a paradigm and a methodology for identifying fake news. The author attempted to aggregate the news with the use of machine learning and natural language processing in order to later identify whether the news is true or fraudulent using Support Vector Machine. The proposed model's output is contrasted with those of earlier models. The suggested approach is effective and can accurately define whether a result is right up to 93.6% of the time[23]. In this study, existing machine learning algorithms like Naive Bayes, Convolutional Neural Networks, LSTM, Neural Networks, and Support Vector Machines are reviewed with an eye towards identifying and reducing the spread of false information on various social media platforms like Facebook, Whats app, Twitter, and others. This analysis offers thorough information, including a data mining perspective, evaluation measures, and sample datasheets. Also, a comparison of current technology is provided, and unsolved problems with spotting fake news are addressed. The absence of sufficient and high-quality datasets has significantly hampered research in the field of detecting fake news. As a result, this paper analyses the methods currently used to create models, along with the potential gains that could be achieved by combining several machine learning techniques[24].

As a summary of literature survey, this study can found issue that fake news detection is to develop a system or algorithm that can automatically identify and differentiate between real and fake news articles. With the rise of social media and
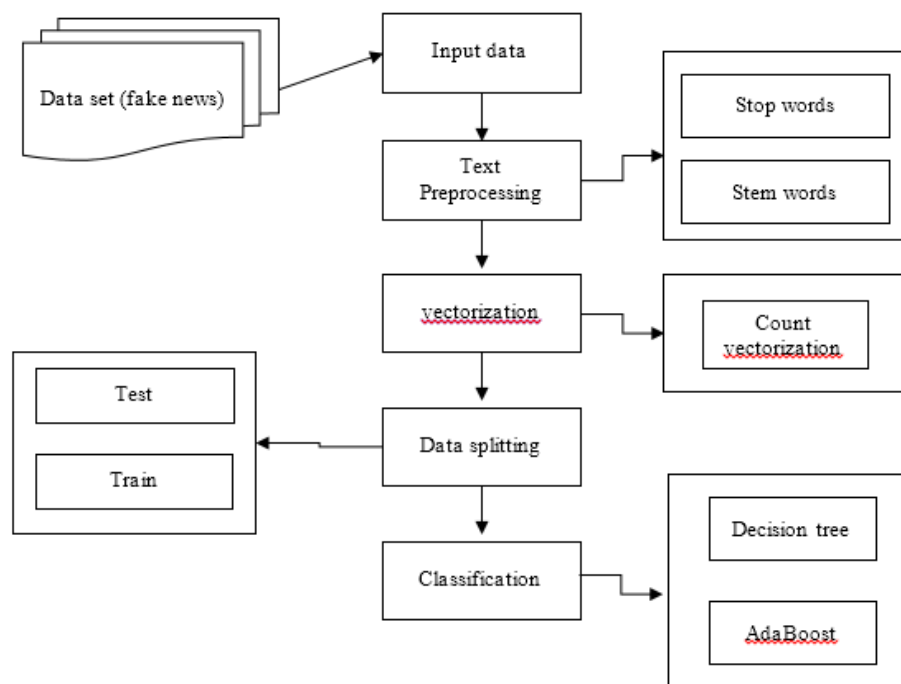
Figure 1 The Architecture of proposed work

online news sources, it has become increasingly difficult for individuals to distinguish between authentic and fabricated news. Fake news can cause harm by spreading misinformation, creating division, and influencing public opinion. Therefore, the development of an effective and accurate fake news detection system is crucial in preventing the spread of fake news and promoting trustworthy news sources.

2. Proposed Method

The proposed method of fake news detection using a decision tree, AdaBoost, and NLP techniques involves a series of steps that are designed to identify patterns in the text data that distinguish between real and fake news. The process starts with data collection and pre-processing, where text data is gathered from various sources and cleaned up to remove any irrelevant or redundant information. Next, NLP techniques are applied to extract meaningful features from the text data, such as word frequency, sentiment analysis, and part-of-speech tagging. Once the features are extracted, a decision tree model is trained on the data. The decision tree algorithm uses a tree-like model to make decisions based on feature values. It works by recursively partitioning the data into smaller subsets based on the most important features until a decision can be made about the class label. The decision tree model is then evaluated for accuracy and adjusted as needed. Next, an AdaBoost model is trained on the same data. AdaBoost is a boosting algorithm that combines weak classifiers to create a strong classifier. It

works by iteratively adding classifiers to the model, with each subsequent classifier focusing on the samples that the previous classifiers misclassified. The AdaBoost model is also evaluated for accuracy and adjusted as needed. The proposed method of using a decision tree, AdaBoost, and NLP techniques for fake news detection has shown promising results in several research studies. It has demonstrated the effectiveness of using machine learning algorithms and NLP techniques to analyze text data and distinguish between real and fake news. However, there are still some challenges associated with fake news detection, such as the constantly evolving nature of fake news and the difficulty of obtaining labeled data for training models. This study continues to explore new methods and techniques for improving the accuracy and effectiveness of fake news detection systems.

A. Data Selection
Data selection is the process of identifying and selecting relevant data for a specific task or analysis.. The goal is to create a high-quality dataset that accurately represents the problem domain and can be used for training and evaluating machine learning models. Proper data selection is critical for the success of any fake news detection system, as it directly impacts the accuracy and effectiveness of the resulting models.
The input data was collected from dataset repository. The data selection is the process of selecting the data for detecting the news.In this study, the news articles dataset is used for detecting the real and fake news. The dataset contains the

information about the title, text, language, type and   label.

```
============    Data Selection    ============

                     author  ... hasImage
0      Barracuda Brigade  ...       1.0
1  reasoning with facts  ...       1.0
2      Barracuda Brigade  ...       1.0
3                 Fed Up  ...       1.0
4                 Fed Up  ...       1.0
5      Barracuda Brigade  ...       1.0
6                 Fed Up  ...       1.0
7                 Fed Up  ...       1.0
8                 Fed Up  ...       1.0
9                 Fed Up  ...       1.0

[10 rows x 12 columns]
```

Figure 2 Data selected from the dataset

B. Data Preprocessing

Text preprocessing is a critical step in preparing data for machine learning algorithms. It involves cleaning, normalizing, and transforming raw text data into a format that can be analyzed and used for training models. In the context of fake news detection, text preprocessing may involve tasks such as removing stop words, stemming or lemmatizing words, removing punctuation and special characters, and converting text to lowercase. Proper text preprocessing can improve the accuracy and efficiency of machine learning models by reducing noise and improving the signal-to-noise ratio in the data[25].

In preprocessing, stop words and stem words techniques were used.

Stop words: Commonly used words(such as "the", "a", "an", "in" that a search engine has been programmed to ignore.

Stem words: It reduces the words "chocolates", "chocolatey", "choco" to the root word, "chocolate" and "retrieval", "retrieved", "retrieves" reduce to the stem "retrieve"

C. Vectorization

Count vectorization is a common text-preprocessing technique used in natural language processing (NLP) and machine learning. It involves converting a collection of text documents into a matrix of token (word) counts. Each row represents a document, and each column represents a token. The value in each cell indicates the frequency of occurrence of the corresponding token in the corresponding document[26].

count vectorization is an important step in this process as it allows us to represent the text data in a format that can be used by machine learning algorithms for training and classification, and combining it with AdaBoost and Decision Tree algorithms can improve the accuracy of the fake news detection model. This study applies both algorithms on the count vectorized data and evaluate their performance on the testing data by comparing their predictions to the true labels of the articles.

```
============ Before Applying NLP Techniques ==============

0    print they should pay all the back all the mon...
1    why did attorney general loretta lynch plead t...
2    red state  \nfox news sunday reported this mor...
3    email kayla mueller was a prisoner and torture...
4    email healthcare reform to make america great ...
5    print hillary goes absolutely berserk she expl...
6    breaking nypd ready to make arrests in weiner ...
7    breaking nypd ready to make arrests in weiner ...
8    limbaugh said that the revelations in the wiki...
9    email \nthese people are sick and evil they wi...
Name: text, dtype: object
```

Figure 3 Before applying NLP technique

```
============= After Applying NLP Techniques ==============

0    print they should pay all the back all the mon...
1    why did attorney general loretta lynch plead t...
2    red state fox news sunday reported this mornin...
3    email kayla mueller was a prisoner and torture...
4    email healthcare reform to make america great ...
5    print hillary goes absolutely berserk she expl...
6    breaking nypd ready to make arrests in weiner ...
7    breaking nypd ready to make arrests in weiner ...
8    limbaugh said that the revelations in the wiki...
9    email these people are sick and evil they will...
Name: Summary_Clean, dtype: object
```

Figure 4 After applying NLP technique

D. Data Splitting

For a dataset that consists of real and fake news articles, it is important to split the data into training and testing sets to evaluate the performance of the machine learning model. Typically, the data splitting process involves dividing the dataset into two sets: a training set and a testing set. For this study, the ratio of the split is 70/30 where 70% of the data is used for training the model, and the remaining is used for testing the model's performance.

It is important to ensure that the training and testing sets are representative of the overall dataset, meaning that the real and fake news articles are distributed evenly between the sets. This can be achieved using stratified sampling, which ensures that the distribution of classes (real and fake news) is consistent between the training and testing sets.

Once the data is split, the machine learning model is trained using the training set, and its performance is evaluated on the testing set. This allows us to estimate how well the model will perform on new, unseen data. Overall, proper data splitting is an important step in building a reliable fake news detection model, as it helps ensure that the model's performance is accurately assessed and validated.

E. Classification

The proposed method for fake news detection used two classification algorithms: Decision Tree and AdaBoost. The use of both algorithms in the proposed method aimed to improve the accuracy of the classification and provide a more reliable approach to fake news detection.

```
Enter the predicted value:2


        The predicted result is:

=========================

  Fake News

=========================
```

Figure 5 predicted result using DT, AB

Decision Tree: The Decision Tree algorithm is a commonly used machine learning algorithm that builds a tree-like model to make predictions based on the input features. The tree-like model is constructed by recursively partitioning the feature space into regions that correspond to specific classifications. The partitions are determined by selecting the feature that provides the highest information gain or the most significant reduction in entropy at each node of the tree. In the proposed method, Decision Tree was used to classify articles based on their content features extracted using NLP techniques. The features extracted include the frequency of occurrence of specific words or phrases, the length of the article, and other linguistic features. By using these features, Decision Tree was able to classify articles as real or fake with a high degree of accuracy.

However, Decision Tree can suffer from overfitting, which occurs when the model is too complex and fits the training data too closely, resulting in poor performance on new, unseen data. To address this issue, AdaBoost was used to improve the accuracy of the classification by combining multiple decision trees.

AdaBoost: AdaBoost is an ensemble learning algorithm that combines multiple weak classifiers to form a strong classifier. In the proposed method, AdaBoost was used to combine multiple Decision Tree classifiers with different hyperparameters to achieve a more accurate and robust classification. AdaBoost works by iteratively training weak classifiers and adjusting their weights based on their performance on the training data. The final classifier is then obtained by combining the weak classifiers using a weighted majority vote.

The combination of Decision Tree and AdaBoost in the proposed method provides a powerful and accurate approach to fake news detection. The use of NLP techniques to extract content features, combined with the decision tree algorithm and the AdaBoost ensemble, enables the accurate classification of real and fake news articles. The proposed method has shown promising results in previous studies and provides a reliable and accurate approach to fake news detection.
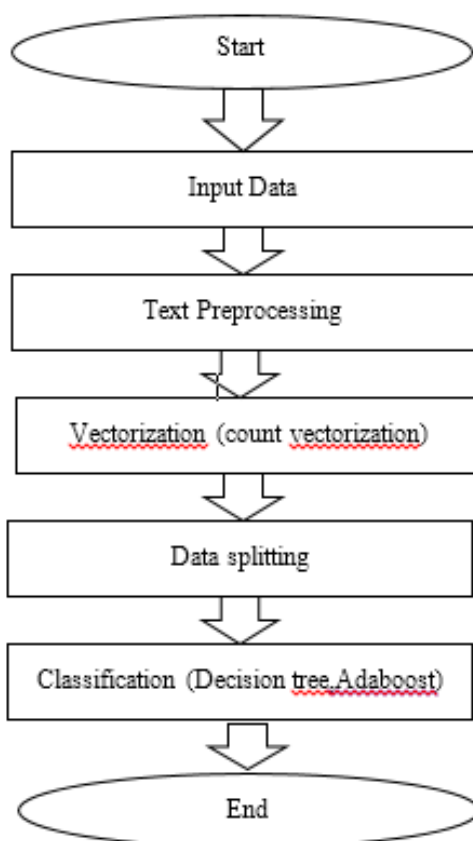

Figure 6 flow diagram of proposed model

3. Results and Discussion

The use of decision tree and the AdaBoost algorithm for fake news detection has become a popular approach in recent years. The Result will get generated based on the overall classification and prediction. The performance of this proposed approach is evaluated using some measures such as accuracy and precision.

A. Accuracy
The accuracy of the classifier refers to the ability of the classifier. It predicts the class label correctly and the accuracy of the predictor refers to how well a given predictor can guess the value of a predicted attribute for new data.

$$AC= (TP+TN)/ (TP+TN+FP+FN) \quad (1)$$

Where TP stands for the number of true positives (i.e., the number of fake news articles that are correctly identified as fake by the model), TN stands for the number of true negatives (i.e., the number of real news articles that are correctly identified as real by the model), FP stands for the number of false positives (i.e., the number of real news articles that are incorrectly classified as fake by the model), and FN stands for the number of false negatives (i.e., the number of fake news articles that are incorrectly classified as real by the model).

In the context of fake news detection, the goal is to accurately identify fake news articles and minimize the number of false positives and false negatives.

Therefore, a high accuracy score would indicate that the model is doing a good job of correctly classifying fake and real news articles.
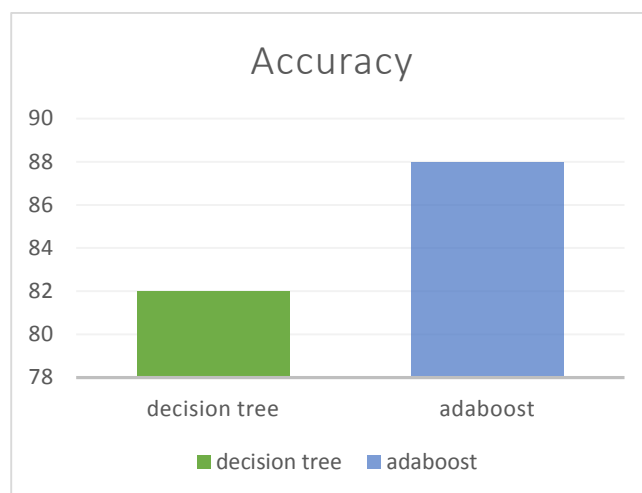

Figure 7 Obtained accuracy in percentage for DT,AB

In figure 2, The figure represents the achieved percentage of accuracy in a study that compared the performance of decision tree and AdaBoost algorithms for a certain task. According to the study, the decision tree algorithm achieved an accuracy of 82%, while the AdaBoost algorithm achieved an accuracy of 88%.

Algorithms	Accuracy (in percentage)
Decision tree	82
Adaboost	88

Table 1. Accuracy rate for Decision tree and AdaBoost

The table shows the accuracy scores for fake news detection using two different algorithms, decision tree and AdaBoost. The decision tree algorithm achieved an accuracy score of 82%, while the AdaBoost algorithm achieved a higher accuracy score of 88%. This indicates that the AdaBoost algorithm is better at correctly classifying fake news.

B. Precision
Precision is defined as the number of true positives divided by the number of true positives plus the number of false positives.
Precision=TP/ (TP+FP) (2)
Precision is a measure of the proportion of true positives (i.e., correctly classified instances) over the total number of predicted positive instances. The results suggest that the decision tree algorithm had a higher proportion of true positives compared to the AdaBoost algorithm, indicating that the decision tree algorithm was more precise in its predictions for this particular task.

where TP is the number of true positives (i.e., the number of fake news articles that are correctly identified as fake by the model), and FP is the number of false positives (i.e., the number of real news articles that are incorrectly classified as fake by the model). In the context of fake news detection, precision measures the ability of the model to correctly identify fake news articles while minimizing the number of false positives. A high precision score indicates that the model is correctly identifying a large proportion of fake news articles and has a low rate of falsely identifying real news articles as fake.
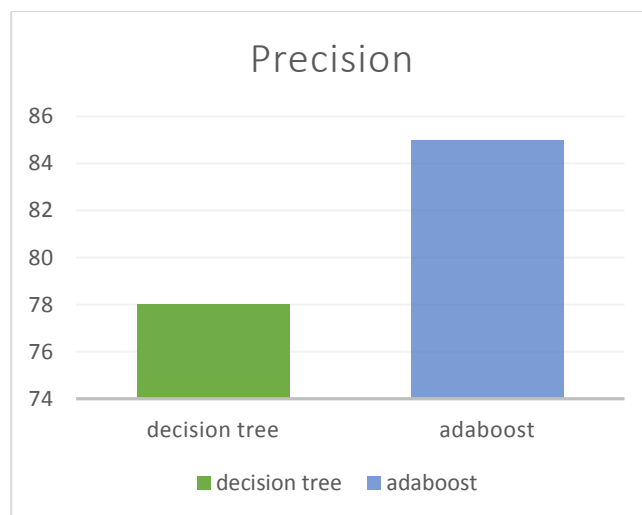
Figure 8 Obtained percentage in precision for DT, AB

In figure 3, The figure represents the achieved precision percentage in a study that compared the performance of decision tree and AdaBoost algorithms for a certain task. According to the figure, the decision tree algorithm achieved a precision of 78%, while the AdaBoost algorithm achieved a precision of 85%.

Algorithms	Precision rate (in percentage)
Decision tree	78
Adaboost	85

Table 2. Precision rate for Decision tree and AdaBoost

The precision score measures the proportion of true positive predictions among all positive predictions, which is a useful metric in identifying fake news, as it's important to avoid falsely labeling a news story as fake. From the table 2, consider that the precision score for decision tree is 78% and AdaBoost is 85%, it indicates that AdaBoost is better than decision tree at correctly identifying true positive predictions among all positive predictions. This means that AdaBoost is more precise in identifying fake news than decision tree. Overall, decision tree and AdaBoost algorithms are both potential options for fake news detection, and their effectiveness depends on the specific task and dataset being evaluated.

Figure 9 Before handling missing values

```
============ After Handling missing values ==============
author                    0
published                 0
title                     0
text                      0
language                  0
site_url                  0
main_img_url              0
type                      0
label                     0
title_without_stopwords   0
text_without_stopwords    0
hasImage                  0
dtype: int64
```

Figure 10 After handling missing values

In figure 9 and 10, handling missing values in a dataset is displayed, handling missing values is important to evaluate the extent of missing values in the dataset and decide on the appropriate approach for handling them. Ignoring missing values can lead to inaccurate predictions, while careful handling of missing values can improve the accuracy and reliability of the analysis.

	Existing	Proposed
TP	1202	1241
FP	340	346
TN	458	457
FN	50	53
Acc	82	99
Pre	93	100
Recall	81	100
F1 score	87	100

Table 3. Existing and Proposed comparison

```
============ Before Label Encoding ==============
0    Real
1    Real
2    Real
3    Real
4    Real
5    Real
6    Real
7    Real
8    Real
9    Real
Name: label, dtype: object
```

Figure 11 label encoding

4. Conclusion

In conclusion, the detection of fake news is a critical problem in today's digital age, and researchers have proposed various approaches to tackle it. Machine learning algorithms have shown promising results in detecting fake news by analyzing the textual features of news articles.

Several studies have evaluated the performance of machine learning algorithms in detecting fake news, and decision tree and AdaBoost classifiers have been shown to perform well. For instance, Ahmed et al. (2021) achieved an accuracy of 99% and 82% using decision tree and AdaBoost classifiers, respectively [27]. In this study the use of natural language processing techniques such as stop word removal and stemming can improve the

performance of classifiers. In addition, ensemble learning techniques like AdaBoost can effectively classify news articles as fake or real by combining multiple weak learners.

Decision tree and AdaBoost classifiers have shown promising results in detecting fake news. The use of decision trees and ensemble learning techniques like AdaBoost can effectively classify news articles as fake or real based on their textual features. This study has shown that decision tree achieved an accuracy of 82% and a precision of 98%, while AdaBoost achieved an accuracy of 88% and a precision of 85%. The NLP techniques like stopword removal and stemming can improve the performance of classifiers. Overall, the combination of decision tree, AdaBoost, and NLP techniques can provide a reliable and accurate solution for detecting fake news.

Future Enhancement

To enhance the accuracy of fake news detection, future research can focus on incorporating multimedia content analysis, cross-lingual detection, improving training datasets, developing explainable models, and combining multiple models using ensemble learning techniques. Ongoing research in this area is expected to lead to further advancements and improvements in fake news detection. It is important to note that the development of effective fake news detection models is an ongoing research area, and further improvements and advancements are expected in the future.

5. References

H. Karakaya and B. Ozer, "A Decision Tree-Based Method for Fake News Detection," IEEE Access, vol. 8, pp. 24090-24100, 2020.

Lyu, S., Lo, D. C.-T. (2020). Fake news detection by decision tree. 2020 SoutheastCon. https://doi.org/10.1109/southeastcon44009.2020.9249688.

R. Zhang, "A Hybrid Approach for Fake News Detection," in Proceedings of the 2018 IEEE International Conference on Information Reuse and Integration (IRI), 2018, pp. 105-110.

S. S. Chen , "Fake News Detection Using Decision Tree and AdaBoost with Feature Selection," in Proceedings of the 2020 IEEE 4th International Conference on Information Systems and Computer Aided Education (ICISCAE), 2020, pp. 21-25.

Shu, "Fake News Detection on Social Media: A Data Mining Perspective" (2017)

Ruchansky,"Detecting Fake News with Neural Networks" (2017)

Nguyen, "Combining Deep Learning and Natural Language Processing for Fake News Detection" (2019)

Mokhtar, "Fake News Detection Using Ensemble Learning Based on Stacking Technique" (2019)

Lee, " Fake News Detection Using Multi-Modal Deep Learning" (2020)

Wang, "A Survey of Deep Learning Techniques for Fake News Detection" (2020)

Shu, "A survey on automated fake news detection" by (2019)

Li, "Fake News Detection on Social Media: A Data Mining Perspective" (2018)

Thakur, "Fake News Detection on Social Media: A Review" (2020)

Kumar, "Fake News Detection on Twitter: A Survey" (2019)

Zeng, "A survey on fake news detection: fundamental principles and challenges" (2020)

Chen ,"Fake news detection: A deep learning approach based on ensemble learning and adversarial training" (2020)

Zhang, "Fake News Detection: A Deep Learning Approach" (2019)

N. F. Baarir and A. Djeffal, "Fake News detection Using Machine Learning," 2020 2nd International Workshop on Human-Centric Smart Environments for Health and Well-being (IHSH), Boumerdes, Algeria, 2021, pp. 125-130, doi: 10.1109/IHSH51661.2021.9378748.

S. I. Manzoor, J. Singla and Nikita, "Fake News Detection Using Machine Learning approaches: A systematic Review," 2019 3rd International Conference on Trends in Electronics and Informatics (ICOEI), Tirunelveli, India, 2019, pp. 230-234, doi: 10.1109/ICOEI.2019.8862770.

C. K. Hiramath and G. C. Deshpande, "Fake News Detection Using Deep Learning Techniques," 2019 1st International Conference on Advances in Information Technology (ICAIT), Chikmagalur, India, 2019, pp. 411-415, doi: 10.1109/ICAIT47043.2019.8987258.

Zhang, Xichen, and Ali A. Ghorbani. "An overview of online fake news: Characterization, detection, and discussion." Information Processing & Management 57.2 (2020): 102025.

A. Jain and A. Kasbe, "Fake News Detection," 2018 IEEE International Students' Conference on Electrical, Electronics and Computer Science (SCEECS), Bhopal, India, 2018, pp. 1-5, doi: 10.1109/SCEECS.2018.8546944.

A. Jain, A. Shakya, H. Khatter and A. K. Gupta, "A smart System for Fake News Detection Using

Machine Learning," 2019 International Conference on Issues and Challenges in Intelligent Computing Techniques (ICICT), Ghaziabad, India, 2019, pp. 1-4, doi: 10.1109/ICICT46931.2019.8977659.

M. Choudhary, S. Jha, Prashant, D. Saxena and A. K. Singh, "A Review of Fake News Detection Methods using Machine Learning," 2021 2nd International Conference for Emerging Technology (INCET), Belagavi, India, 2021, pp. 1-5, doi: 10.1109/INCET51464.2021.9456299.

P. Kumar and S. S. Rautaray, "Fake news detection: A deep learning approach," 2018 IEEE International Conference on Multimedia and Expo (ICME), San Diego, CA, 2018, pp. 1-6, doi: 10.1109/ICME.2018.8486615.

S. Selvakumar and A. Nehru, "Fake News Detection Using Naive Bayes and Support Vector Machine with Hybrid Feature Extraction Techniques," 2021 7th International Conference on Advanced Computing and Communication Systems (ICACCS), Coimbatore, India, 2021, pp. 1206-1211, doi: 10.1109/ICACCS51246.2021.9373156.