International Conference on Innovative Data Communication Technologies and Application (ICIDCA-2023)
IEEE Xplore Part Number: CFP23CR5-ART; ISBN: 979-8-3503-9720-8

Fake News Detection using NLP

Mohammed Ali Shaik
Department of Computer Science &Artificial Intelligence
SR University
Warangal, Telangana State, India
niharali@gmail.com

Makkaji Yasha Sree
Department of Computer Science & Engineering
SR University
Warangal, Telangana State, India
19k41a05g2@sru.edu.in

Sanka Sri Vyshnavi
Department of Computer Science & Engineering
SR University
Warangal, Telangana State, India
19k41a05g8@sru.edu.in

Thogiti Ganesh
Department of Computer Science & Engineering
SR University
Warangal, Telangana State, India
19k41a05e7@sru.edu.in

Dasari Sushmitha
Department of Computer Science & Engineering
SR University
Warangal, Telangana State, India
19k41a05c5@sru.edu.in

Narmetta Shreya
Department of Computer Science & Engineering
SR University
Warangal, Telangana State, India
19k41a05g5@sru.edu.in

2023 International Conference on Innovative Data Communication Technologies and Application (ICIDCA) | 979-8-3503-9720-8/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICIDCA56705.2023.10100305

***Abstract*—In the age of digital media, fake news is a serious problem because it spreads misinformation and harms individuals, organizations, and even entire nations which is a challenging aspect. This study proposes a machine learning approach for detecting fake news. In the proposed approach, a categorization model is developed with four different types of machine learning algorithms, evaluating the content and aesthetic components of news stories. The performance of the proposed model is analyzed by using a large dataset of real and fake news articles and the results show that it outperforms many existing systems. The proposed findings demonstrate the potential of machine learning techniques, such as logistic regression, decision tree, random forest, and passive aggressive algorithms to address the fake news detection challenges.**

Keywords—Machine learning, Data, prediction model, Classification, Logistic regression, Random Forest, Decision tree and Passive Aggressive, Fake news detection.

I. INTRODUCTION

The deliberate spread of incorrect or misleading information through different media is referred to as fake news, also known as disinformation. Fake news has become a widespread issue with the rapid rise of the internet and social media, and it now poses a threat to society in many ways, including by inciting fear and distrust, influencing public opinion and decision-making, and even producing political instability. Therefore, it has become crucial for governments, media outlets, and individuals to identify and stop the spread of fake news.

To perform fake news detection, this study intends to create a system that can recognize false news articles with accuracy. To accomplish this, we will examine the content of news items to establish their veracity using machine learning algorithms and methods of natural language processing [30].

The system will be trained on a large dataset of news articles labeled as real or fake, and it will extract features such as the type of language used, the presence of certain keywords, and the sentiment expressed in the text. The machine learning model will then use these features to make a prediction about the authenticity of the news article.

The final product will be a reliable and accurate fake news identification system that can aid in halting the spread of false information and encouraging the spread of true information. The system will be assessed against other current techniques for fake news identification using common measures including accuracy, precision, recall, and F1 score. By helping to create a solution to the fake news problem, this paper has the potential to have a big influence.

II. LITERATURE REVIEW

Paper– [1]: "Combining Textual and Network Features for Fake News Detection on Social Media" by S. S. Alqahtani, M. Alshomrani, and A. Alshomrani. In this study, the authors used the passive aggressive algorithm to classify news articles as real or fake, based on a combination of textual and network features. The study was published in 2021. The authors evaluated their method on a dataset of real and fake news articles collected from various sources on the web. They found that the PA algorithm combined with textual and network features outperformed several other methods for fake news detection [12].

Merits:The PA algorithm is simple and fast, making it well-suited to the problem of "fake news detection on social media".By combining textual and network features, the authors were able to improve the performance of the PA algorithm for fake news detection [13].

Demerits:The algorithm may not perform well when the data is noisy or highly unstructured, as is often the case in social media platforms [14].

Paper– [2]:"An Approach for Fake News Detection using Passive Aggressive Algorithm on Social Media" by H. R. Nandini and H. B. Kavyashree. In this study, the authors classified news pieces on social media as real or fake using the passive aggressive algorithm based on characteristics such the story's source, the presence of keywords, and the attitude the text conveyed. In 2020, the study was released. On a dataset of news pieces gathered from several social media networks, the authors tested their methodology [15]. They discovered that the PA algorithm beat a number of other techniques, including Naive Bayes and Decision Tree algorithms, for the detection of bogus news [16].

Merits:By incorporating a variety of features, the authors were able to improve the performance of the PA algorithm for fake news detection.

Demerits:The algorithm is sensitive to the choice of hyperparameters, and its performance can degrade if the hyperparameters are not set appropriately.

Paper– [3]: In "Fake News Detection using Random Forest with Sentiment Analysis" by B. K. Singh and S. Jain. In this study, the authors used the random forest algorithm to classify news articles as real or fake, based on features such as the source of the news, the presence of keywords, and the sentiment expressed in the text. The study was published in 2021. The authors evaluated their method on a dataset of news articles collected from various sources on the web [17]. They found that the RF algorithm combined with sentiment analysis outperformed several other methods for fake news detection, including Naive Bayes and Logistic Regression algorithms [18].

Merits:The RF algorithm is robust and can handle complex data distributions and non-linear relationships between features and the target variable.

Demerits:The RF algorithm is computationally intensive and requires a lot of memory, making it less suitable for real-time detection of fake news on social media platforms.

Paper– [4]: "A Random Forest Based Approach for Fake News Detection in Social Media" by M. J. Aslam and A. S. F. Zaidi. In this study, the authors used the random forest algorithm to classify news articles on social media as real or fake, based on features such as the source of the news, the presence of keywords, and the sentiment expressed in the text. The study was published in 2019 [19]. The authors evaluated their method on a dataset of news articles collected from various social media platforms [20]. They found that the RF algorithm outperformed several other methods for fake news detection, including Naive Bayes and Logistic Regression algorithms [21].

Merits:By incorporating a variety of features, the authors were able to improve the performance of the RF algorithm for fake news detection.

Demerits:The algorithm can be sensitive to overfitting, especially when the data is highly unstructured or noisy.

Paper– [5]:"Combating misinformation in social media with machine learning: a survey" by Nikolaos Aletras, ArkaitzZubiaga, and David Corney (2017). The authors provide an overview of the various ML algorithms used for fake news detection, including logistic regression. They also discuss the challenges and future directions of the field, including the need for large annotated datasets and the development of robust evaluation metrics [22].

Merits:Using logistic regression in this context include its simplicity, interpretability, and the ability to handle large datasets efficiently.

Demerits:Logistic Regression isnot suitable for more complex problems where the relationship between the features and target is not linear.

Paper– [6]: In "Fake News Detection on Social Media: A Data Mining Perspective" by Arjun Mukherjee, Dmitry Davidov, and Eugene Agichtein. This study used logistic regression to classify fake news articles based on features such as sentiment, subjectivity, and credibility of the source.

Merits:Logistic regression can handle large datasets, making it well-suited to the problem of fake news detection on social media.

Demerits:Logistic regression may not perform well when the data is highly imbalanced, such as in the case of fake news detection, where the proportion of fake news is small relative to the amount of real news.

Paper– [7]: "Fake News Detection Using Decision Trees and Naive Bayes" by J. Chen and J. Liu. In this study, the authors classified news stories as real or fake using decision trees and Naive Bayes algorithms based on characteristics including the news source, the existence of terms, as well as the emotion conveyed in the text. In 2020, the study was released [23]. On a dataset of news stories gathered from diverse online sources, the authors tested their strategies. They discovered that the DT and Naive Bayes algorithm combo beat a number of other techniques for identifying fake news, including Logistic Regression and Random Forest algorithms [24].

Merits:DT algorithms are capable of handling complex relationships between features and target variables.

Demerits:DT algorithms are sensitive to small changes in the training data, making them unstable

Paper– [8]:"Fake News Detection Using Decision Trees and Random Forest" by Y. Zhang and L. Wang. In this study, the investigators classified news stories as legitimate or fraudulent using decision trees and random forest algorithms. 2019 saw the publication of the study. A dataset of news stories compiled from multiple online sources served as the basis for the authors' approach evaluation [25]. They discovered that the DT and random forest algorithms beat a number of other techniques for identifying bogus news, including Logistic Regression and Naive Bayes algorithms [26].

Merits:Random Forest algorithms are robust to overfitting, making them a popular choice for many classification tasks.

Demerits:DT algorithms can be prone to overfitting if not properly tuned.

III. PROBLEM DEFINITION

The task is to accurately classify each news article as either real or fake given a set of news articles. The difficulty in defining what constitutes "fake news," as well as the difficulty in automatically detecting such news, is at the heart of this problem.

Misinformation, propaganda, satire, and even conspiracy theories are all examples of fake news. It can be disseminated via a variety of media, including traditional news outlets, social media platforms, and even personal websites. Fake news articles' content can also be designed to appeal to emotions, biases, or beliefs, making them difficult to distinguish from legitimate news.

Furthermore, it has become challenging for people to distinguish between true and fake news due to the quick dissemination of false information online. This is particularly problematic in politics because false information has the power to influence people's opinions and actions.

As a result, both from a sociological and a technical perspective, the issue of identifying fake news is urgent. Advanced natural language processing methods and machine learning algorithms must be combined in order to effectively discern between authentic and false news stories.

A. DATASET

For this paper we used two data sets obtained from Kaggle. 1) News.csv 2) Fake-Real The dataset consists of 12 attributes [27]

The first dataset contains three attributes which are

- Title
- Text
- Label- (Fake or Real)

[7] Dataset1.head()

	Unnamed: 0	title	text	label
0	8476	You Can Smell Hillary's Fear	Daniel Greenfield, a Shillman Journalism Fello...	FAKE
1	10294	Watch The Exact Moment Paul Ryan Committed Pol...	Google Pinterest Digg Linkedin Reddit Stumbleu...	FAKE
2	3608	Kerry to go to Paris in gesture of sympathy	U.S. Secretary of State John F. Kerry said Mon...	REAL
3	10142	Bernie supporters on Twitter erupt in anger ag...	— Kaydee King (@KaydeeKing) November 9, 2016 T...	FAKE
4	875	The Battle of New York: Why This Primary Matters	It's primary day in New York and front-runners...	REAL

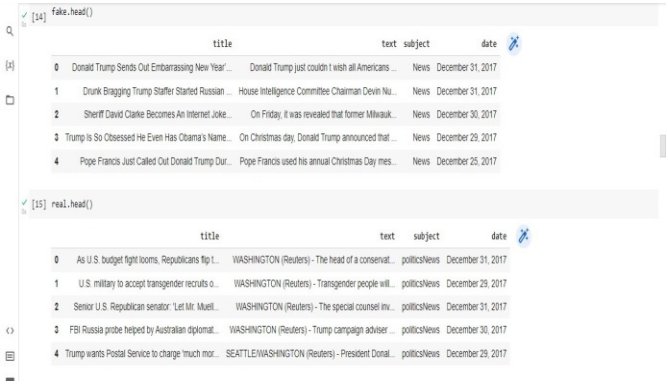
Fig. 1. Dataset attribute description

The dataset contains 6335 rows of data with three columns

The second dataset contains four attributes which are

- Title
- Text
- Subject
- Date

The dataset contains 21417 rows of Real data and 23481 rows of fake data with four columns each. Real data csv file contains Political News and World News, Fake data csv file contains Political News, Left News, Govt News, Us News, Middle-East News. By concatenating both real and fake news now we have around 44k rows with four columns.



[14] fake.head()

	title	text	subject	date
0	Donald Trump Sends Out Embarrassing New Year'...	Donald Trump just couldn t wish all Americans ...	News	December 31, 2017
1	Drunk Bragging Trump Staffer Started Russian ...	House Intelligence Committee Chairman Devin Nu...	News	December 31, 2017
2	Sheriff David Clarke Becomes An Internet Joke...	On Friday, it was revealed that former Milwauk...	News	December 30, 2017
3	Trump Is So Obsessed He Even Has Obama's Name...	On Christmas day, Donald Trump announced that ...	News	December 29, 2017
4	Pope Francis Just Called Out Donald Trump Dur...	Pope Francis used his annual Christmas Day mes...	News	December 25, 2017

[15] real.head()

	title	text	subject	date
0	As U.S. budget fight looms, Republicans flip t...	WASHINGTON (Reuters) - The head of a conservat...	politicsNews	December 31, 2017
1	U.S. military to accept transgender recruits o...	WASHINGTON (Reuters) - Transgender people will...	politicsNews	December 29, 2017
2	Senior U.S. Republican senator: 'Let Mr. Muell...	WASHINGTON (Reuters) - The special counsel inv...	politicsNews	December 31, 2017
3	FBI Russia probe helped by Australian diplomat...	WASHINGTON (Reuters) - Trump campaign adviser ...	politicsNews	December 30, 2017
4	Trump wants Postal Service to charge 'much mor...	SEATTLE/WASHINGTON (Reuters) - President Donal...	politicsNews	December 29, 2017

Fig. 2. Cleaned dataset description

The graph below shows different types of news such Political News, Left News, Govt News, Us News, Middle-East News.

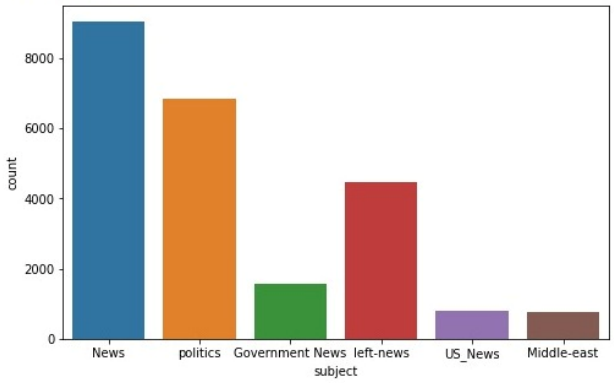


Fig. 3. Classification of data attributes

The graph below shows the different types of news that real.csv file contains such as Political News and World News

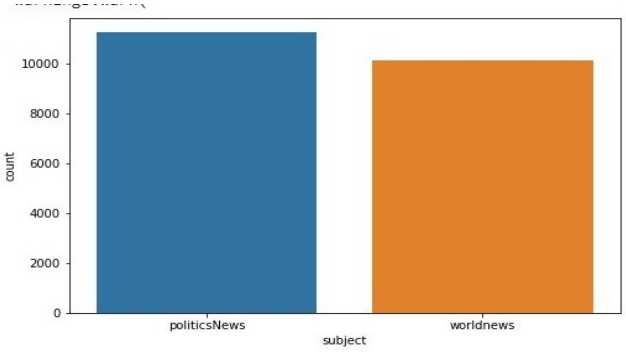


Fig. 4. Types of data

B. DATA PRE-PROCESSING

Data preprocessing is an important step in the fake news detection process as it helps to prepare the data for further analysis and modeling. The following are the steps involved in data preprocessing for fake news detection:

1) *Data Gathering*: The first stage is to gather pertinent news articles and other important data, such as the date of publication, the topic, and the headline.

2) *Data Cleaning*: The gathered data must then be cleaned by eliminating extraneous information, fixing mistakes, and handling missing numbers. This can be achieved by employing strategies like deleting stop words, lowercasing all text, and removing special characters.

3) *Text normalization*: It is a process of putting text into a format that is generally accepted. This can be achieved by deleting numerals, stemming words, and changing all text to lowercase.

4) *Text tokenization*:It is the method of disassembling a statement into its constituent words. Tools like the Natural Language Toolkit (NLTK) or regular expressions can be used for this.

5)*Feature Engineering*: In order to improve the data representation for the false news detection model, new features are created from the existing data.

6) *Text vectorization*: It is the process of transforming text into numerical data that may be fed into a machine learning model. Techniques like "bag-of-words" analysis and term frequency-inverse document frequency can be used for this (TF-IDF).

7)*Split Data into Training and Testing Sets*: The preprocessed data must be divided into training and testing sets in order to properly train and test the false news detection model.

A word cloud is a graphic depiction of the words that appear most frequently in a text or group of texts. The words are arranged into a cloud-like pattern, with the size of each word corresponding to how frequently it appears in the text. Less frequently used terms are shown in smaller font sizes, while the most often used words are presented in bigger font sizes.Word clouds are frequently used for summarising and presenting vast volumes of text data in text analysis and data visualisation. They can be helpful for highlighting words and phrases which are often used as well as for rapidly recognising the most crucial ideas or topics in a document. To enhance visual interest as well as convey more information, word clouds can also be altered with different colours, forms, and font styles.

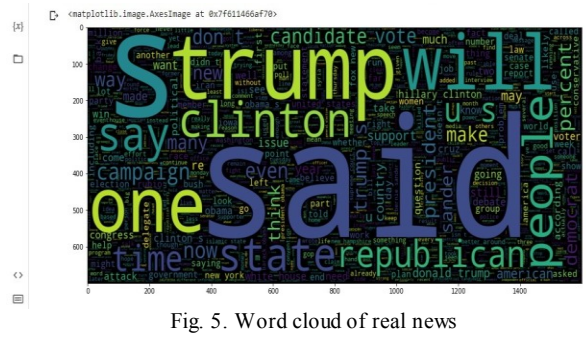


Fig. 5. Word cloud of real news

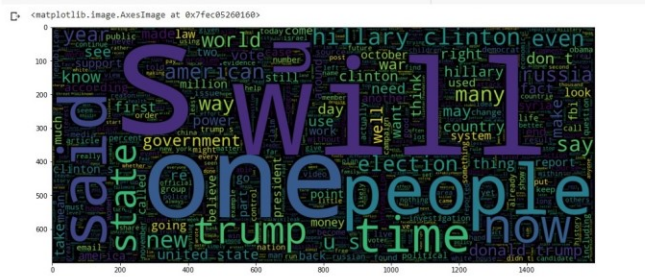


Fig. 6. Word cloud of fake news

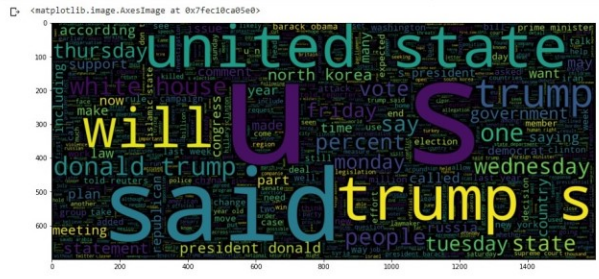


Fig. 7. Word cloud of real news of dataset-2

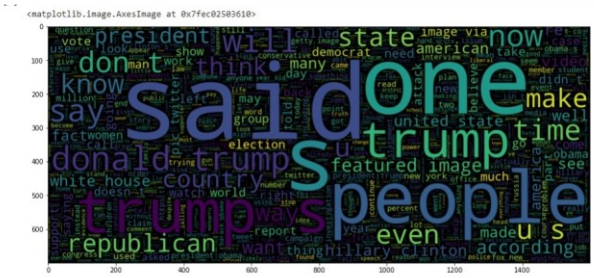


Fig. 8. Word cloud of fake news of dataset-2

C. *ALGORITHMS*

We use a variety of Machine Learning models, all of which are different Regression models, to solve our Prediction problem and the top four are:

- Decision Tree [28]
- Random Forest [29]
- Passive Aggressive [31]
- Logistic Regression [27].

So let's prepare our data for our machine learning model's training and testing.

- *Decision Tree*: Building a coaching model that can be utilized to predict the categorization or cost of goal variables by mastering choice policies drawn from training data is the main purpose of the decision tree algorithm, which is a subset of the supervised learning algorithm family. Regression and classification challenges can be resolved using the decision tree technique. The Decision Tree is mostly utilised for grouping purposes. Additionally, a common categorization model in data mining is the decision tree. Every tree is made up of nodes and branches. Each node represents a class of elements that need to be categorised, and each subset specifies a price that the node can accept. Selection bushes have established several implementation fields due to their simple evaluation and accuracy on a few information forms. Decision tree classifiers are praised for providing an excellent perspective on performance results. Optimized splitting parameters and better tree pruning techniques (ID3 [18], C4.5 [19], CART [20], CHAID [21], and QUEST [22]) are frequently employed by all known information classifiers due to their high precision. The distinct datasets are used to extract training samples from a huge record set, which has an impact on the test set's precision.
- *Random Forest*:In order to import a previously trained version of the network used for having to implement training over thousands of Laptops data, Random Forest, an ensemble of decision trees, uses a "Laptops database." As a result, it will build up a library of additional features that denotes accuracy of 87% and r2 score is 0.15%, which are best when compared to other algorithms.
- *Passive Aggressive*: For classification and regression issues, the Passive Aggressive (PA) algorithm is an online machine learning technique. The technique is intended to be quick and effective, making it

appropriate for real-time applications and large-scale datasets. A linear classifier or regressor is incrementally adjusted with each training example in the PA method. The discrepancy between the projected label or value as well as the actual label or value determines the update step. The update step is estimated to have a limited effect on the model's prior predictions while yet having a high degree of confidence in the prediction for the current example. The PA algorithm's capability to manage data instances with significant prediction errors or incorrect classifications, which might happen often in real-world applications, is one of its important strengths. In these circumstances, the algorithm is intended to be passive for situations that are correctly classified while also being more aggressive in rectifying the forecast error.

- *Logistic Regression:* A statistical technique for analysing any dataset in which one or more predictor variables affect a result is called logistic regression. It is applied to categorical outcomes or dependent variables in classification issues.A logistic function, which generates a probability between 0 and 1, is used in logistic regression to represent the connection between the independent factors and the dependent variable. Given the values of the various independent variables, the logistic function simulates the likelihood that the dependent variable (such as class membership) would take on a specific value.Due to its ease of use and interpretability, logistic regression is a common machine-learning technique. It is very simple to use and can handle interactions between the independent as well as dependent variables that are both linear and non-linear. Nevertheless, it can only be applied to situations involving binary classification or multiple classes when many models are trained then integrated.

D. BUILDING THE MODEL

- *Defining the Problem*: The goal of building a fake news detection model is to determine whether the news is authentic or fake. This is an important issue because fake news has the potential to harm individuals and society by spreading misinformation and influencing public opinion.
- *Preparing the Data*: The next step is to pre-process the data in order to prepare it. the gathering of pertinent news articles, data cleaning, text normalisation, text tokenization, and development of new features. We have gathered two distinct datasets with various title, text, subject, date, and label properties. Then, designating it as an article, consolidated the title as well as text into such a single column. We removed that column because it wasn't really useful given the article's publication date. The column labels that indicate if whether news is fake or real have been modified to read "0" for fake news and "1" for true news. We prepared the data in this way.
- *Selecting a Model*: We have identified four key models to train using the data out of the many machine learning methods that can be utilised for false news identification. These include the algorithms Decision Tree, Logistic Regression, Random Forest, and Passive Aggressive.
- *Model Training*: Using the training set of data, we successfully trained the model. In order to do this, pre-processed data must be fed to the model in order for it to recognize patterns.
- *Evaluating the Model*: Using the testing data, we compared the projected results to the actual results and calculated accuracy to assess the model's performance.

IV. RESULTS AND COMPARITIVE STUDY

- *Passive Aggressive*: It is a linear classifier algorithm and can be used for binary or multiclass classification problems. It is known for its fast-training time and ability to handle large data sets efficiently. It's mainly used for online learning, where new data can continuously be added to the model and updated. Its main weakness is that it can be sensitive to outliers and irrelevant features, which can negatively impact its performance.

We achieved 97.86% accuracy using the PASSIVE AGGRESSIVE algorithm.

- *Decision Tree*: It is a straightforward but effective approach that works both for regression and classification issues. It is perfect for describing outcomes to stakeholders who really are unfamiliar with technical nuances because it is simple to interpret and visualise. It is simple to manage non-linear correlations between features and goal variables since the algorithm divides the data into progressively lower subsets based on the characteristics. So, when tree is allowed to expand too deep, it is particularly prone to overfitting, which can result in subpar performance on unobserved data.

We achieved 95.29% accuracy using the DECISION TREE algorithm.

- *Logistic Regression*: For problems involving binary and multiple classes in classification, it is a linear algorithm. With a minimal to medium-sized data sets, it is quick to train and effective. It is simple to understand and offers information about how features relate to the desired variable. The assumption that characteristics and target variables have a linear relationship, however, may not necessarily be true in real-world data.

We achieved 96.65% accuracy using the LOGISTIC REGRESSION algorithm.

- *Random Forest*: To produce predictions based on many decision trees, data mining and machine learning use the ensemble learning technique known as random forest. Several decision trees are trained using randomly chosen data subsets, and then their predictions are combined using weighted average or majority voting. Especially in comparison to decision trees, it is significantly accurate and much less prone to overfitting, although it requires more computing.

We achieved 95.81% accuracy using the RANDOM FOREST algorithm.

A. PREDICTIONS

ALGORITHM	ACCURACY
Random Forest [12]	95.81%
Passive Aggressive [13]	97.86%
Logistic Regression [14]	96.65%
Decision Tree [15]	95.29%

We have used these models, and the accuracies for these models are

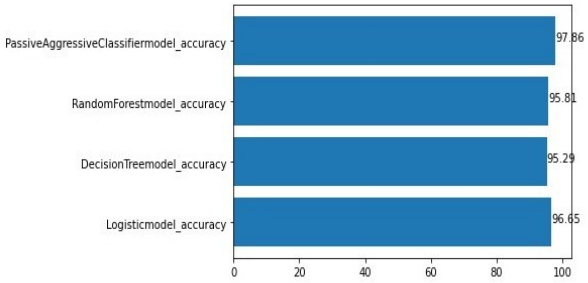


Fig. 9. Obtained Accuracies

B. ENVIRONMENT

Google Colab is a free cloud-based platform that provides access to powerful computing resources and a Jupyter notebook environment for data scientists and machine learning engineers. The platform is designed to allow users to collaborate and share their work, making it an ideal environment for conducting machine learning applications, including fake news detection.

In Google Colab, users have access to GPUs and TPUs, which can greatly speed up the training of machine learning models. This is particularly useful for large and complex models that would otherwise require a lot of computational resources and time to train. With Google Colab, users can start training their models in minutes, without having to worry about setting up their own hardware or software environment.

The Jupyter notebook environment in Google Colab provides a convenient and interactive way to write, execute, and visualize code. Users can write and run their code in the browser, without having to install any software or dependencies on their local machine. The notebooks can be easily shared with others, making it easy to collaborate with team members or share the results with a wider audience.

In the context of a fake news detection paper, Google Colab can be used to train and evaluate machine learning models on large datasets, and to perform data preprocessing and feature extraction. The Jupyter notebooks can be used to document the steps taken in this paper, to record the results, and to share the findings with others.

V. CONCLUSION

To summarize, detecting fake news is a complex problem that necessitates a multidisciplinary approach. Collecting and preprocessing data, selecting and training machine learning algorithms, and fine-tuning the models for improved performance are all steps in the development of effective fake news detection models. The quality and quantity of data, as well as the algorithms and features used, all have a significant impact on the performance of fake news detection models.

Despite these obstacles, fake news detection models have the potential to make a significant contribution to the fight against misinformation. These models can help to mitigate the spread of false information and protect individuals and society from its harmful effects by automatically identifying and flagging it. Furthermore, as technology advances and machine learning algorithms become more sophisticated, fake news detection models are expected to become even more effective in detecting and combating fake news.

REFERENCES

[1] N. R. de Oliveira, D. S. V. Medeiros, and D. M. F. Mattos, ‘‘A sensitivestylistic approach to identify fake news on social networking,’’ IEEE SignalProcess. Lett., vol. 27, pp. 1250–1254, 2020.

[2] A. R. Merryton and G. Augasta, ‘‘A survey on recent advances in machinelearning techniques for fake news detection,’’ Test Eng. Manag, vol. 83,pp. 11572–11582, 2020.

[3] M. Mahyoob, J. Algaraady, and M. Alrahaili, ‘‘Linguistic-based detectionof fake news in social media,’’ Int. J. English Linguistics, vol. 11, no. 1,p. 99, Nov. 2020Journal of Advanced Research in Computer and Communication Engineering ISO 3297:2007 Certified Vol. 6, Issue 12, December 2017

[4] N. R. de Oliveira, P. S. Pisa, M. A. Lopez, D. S. V. de Medeiros, andD. M. F. Mattos, ‘‘Identifying fake news on social networks based onnatural language processing: Trends and challenges,’’ Information, vol. 12,no. 1, p. 38, Jan. 2021.

[5] D. D. N. Caragea, D. M. Caragea and K. C. Al-Kofahi, "A Machine Learning-based System for Detecting Fake News in Social Media," in Proceedings of the 56th Hawaii International Conference on System Sciences, 2019, pp. 3220-3229.

[6] H. Farid and K. Abdullah, "Fake News Detection Based on Machine Learning Algorithms," in 2019 International Conference on Information Technology, Islamic Republic of Iran, 2019, pp. 1-6.

[7] S. R. S. S. Prabaharan and R. Priyanka, "Fake News Detection using Machine Learning Algorithms," in 2020 International Conference on Advances in Computing, Communications and Informatics (ICACCI), 2020, pp. 1190-1195.

[8] J. Kim and J. Kim, "Fake News Detection Using Multi-Source Information and Machine Learning Techniques," in Proceedings of the 2019 International Conference on Computational Science and Computational Intelligence, 2019, pp. 329-334.

[9] Amjad M, Sidorov G, Zhila A, Gómez-Adorno H, Voronkov I, Gelbukh A. 2020. Bend the truth: benchmark dataset for fake news detection in Urdu language and its evaluation. Journal of Intelligent & Fuzzy SystemsRegression Analysis,” vol. 6, no. 5.

[10] Mohammed Ali Shaik, Praveen Pappula, T Sampath Kumar, Predicting Hypothyroid Disease using Ensemble Models through Machine Learning Approach, European Journal of Molecular & Clinical Medicine, , 9(7), 6738-6745, (2022).

[11] M. A. Shaik, S. k. Koppula, M. Rafiuddin and B. S. Preethi, COVID-19 Detector Using Deep Learning, International Conference on Applied Artificial Intelligence and Computing (ICAAIC), 443-449,(2022).

[12] T.Sampath Kumar, B.Manjula, Security Issue Analysis on Cloud Computing Based System,International Journal of Future Generation Communication and Networking, 12(5),143 – 150, (2019).

[13] Mohammed Ali Shaik and Dhanraj Verma, Prediction of Heart Disease using Swarm Intelligence based Machine Learning Algorithms, International Conference on Research in Sciences, Engineering &Technology, Published by AIP Publishing. 978-0-7354-4368-6, 2022, pp. 020025-1 to 020025-9.

[14] YerrollaChanti, Bandi Bhaskar, NagendarYamsani, “Li-Fi Technology Utilized In Leveraged To Power In Aviation System Entertainment Through Wireless Communication”, J. Mech. Cont.& Math. Sci., Vol.-15, No.-6, June (2020) pp 405-412.

[15] Mohammed Ali Shaik and Dhanraj Verma, "Predicting Present Day Mobile Phone Sales using Time Series based Hybrid Prediction

Model", International Conference on Research in Sciences, Engineering & Technology, AIP Conf. Proc. 2418, (2022) pp. 020073-1 to 020073-9

[16] Mohammed Ali Shaik, MD. Riyaz Ahmed, M. Sai Ram and G. Ranadheer Reddy, "Imposing Security in the Video Surveillance", International Conference on Research in Sciences, Engineering & Technology, AIP Conf. Proc. 2418, 020012-1–020012-8, (2022), pp. 020012-1 to 020012-8.

[17] Mohammed Ali Shaik, Geetha Manoharan, B Prashanth, NuneAkhil, Anumandla Akash and Thudi Raja Shekhar Reddy, "Prediction of Crop Yield using Machine Learning", International Conference on Research in Sciences, Engineering & Technology, AIP Conf. Proc. 2418, (2022), pp. 020072-1–020072-8,

[18] Mohammed Ali Shaik and Dhanraj Verma, Enhanced ANN training model to smooth and time series forecast, IOP Conf. Ser.: Mater. Sci. Eng. 981, (2020), pp. 022038

[19] T. Sampath Kumar, B. Manjula, Mohammed Ali Shaik, P. Praveen, A Comprehensive Study on Single Sign on Technique International Journal of Advanced Science and Technology (IJAST), 127, 156-162, (2019)

[20] Mohammed Ali Shaik, Dhanraj Verma, P Praveen, K Ranganath and Bonthala Prabhanjan Yadav, RNN based prediction of spatiotemporal data mining, IOP Conf. Ser.: Mater. Sci. Eng. 981, (2020), pp.022027.

[21] T. Sampath Kumar, B. Manjula, D. Srinivas, A New Technique to Secure Data Over Cloud Jour of Adv Research in Dynamical & Control Systems, 11, (2017), pp 145-149.

[22] Mohammed Ali Shaik and Dhanraj Verma, Deep learning time series to forecast COVID-19 active cases in INDIA: A comparative study, IOP Conf. Ser.: Mater.Sci.Eng.981, (2020), pp. 022041.

[23] P Praveen, M Ranjith Kumar, Mohammed Ali Shaik, R RaviKumar and R Kiran, The comparative study on agglomerative hierarchical clustering usingnumerical data, IOP Conf. Ser.: Mater. Sci. Eng. 981, 2020, pp. 022071.

[24] P. Praveen, B.Rama, An Efficient Smart Search Using R Tree on Spatial Data Journal of Advanced Research in Dynamical and Control Systems, 4, (2019), pp.1943-1949.

[25] D Kothandaraman, N Praveena, K Varadarajkumar, B Madhav Rao, Dharmesh Dhabliya, Shivaprasad Satla, Worku Abera, Intelligent Forecasting of Air Quality and Pollution Prediction Using Machine Learning, Adsorption Science & Technology, 2022, (2022).

[26] Sallauddin Mohmmad, Ramesh Dadi, D Kothandaraman, E Sudarshan, Syed Nawaz Pasha, Mohammed Ali Shaik, A survey machine learning based object detections in an image, AIP Conference Proceedings,2418, (1), (2022), pp. 020024.

[27] A Balasundaram, D Kothandaraman, S Ashokkumar, E Sudarshan, Chest X-ray image based COVID prediction using machine learning, AIP Conference Proceedings, 2418(1), 2022, pp. 020079.

[28] Mohammed Ali Shaik, Time Series Forecasting using Vector quantization", International Journal of Advanced Science and Technology (IJAST), 29(4), 169-175, (2020).

[29] Mohammed Ali Shaik, T. Sampath Kumar, P. Praveen, R. Vijayaprakash, Research on Multi-Agent Experiment in Clustering", International Journal of RecentTechnology and Engineering (IJRTE), 8(1S4), 1126-1129, (2019).

[30] Mohammed Ali Shaik, A Survey on Text Classification methods throughMachine Learning Methods", International Journal of Control and Automation (IJCA),12(6), 390-396, (2019).

[31] R. Ravi Kumar, M. Babu Reddy and P. Praveen, A review of feature subset selection on unsupervised learning Third International Conference on Advances in Electrical, Electronics, Information, Communication and Bio-Informatics (AEEICB), (2017), pp.163-167.