

# Detection of Covid-19 Fake News text data using Random Forest and Decision tree Classifiers

Abdullah Yahya Abdullah Amer<sup>1</sup> and Tamanna Siddiqui<sup>2</sup>

<sup>1&2</sup>Department of Computer Science, Aligarh Muslim University

Aligarh, Uttar Pradesh, India

<sup>1</sup>abdullahamer881@gmail.com, <sup>2</sup>ja\_zu\_siddiqui@hotmail.com

**Abstract:** In current days, Fake news has been growing with significant numbers for numerous political, business, and social reasons. Such news has also become ubiquitous strongly in the online world or social network. People/virtual communities can get contaminated quickly by such fake news. Such news related to Coronavirus covid-19 has dramatically influenced the offline community also. So, there is required a great from self-awareness and community for and better understanding covid-19 informational change the fake news that people spread for political or economic purposes amid this the Corona pandemic crisis to spread terror in society, without taking considerate the feelings of the public. In research articles, particularly these days within covid-19 informational change, we are interested in knowing the algorithms of text mining and machine learning that have the ability to handle and distinguish among real and fake news for Coronavirus covid-19. Through applied two machine learning supervised algorithms, i.e., Random forest and decision tree classifiers to detect Coronavirus covid-19 fake news with our model, Count Vectorizer and Document Frequency Vectorizer as feature extraction after making a set of the initial set such as preprocess and normalization of the dataset. Our proposed detection and the algorithm's ability to differentiate and verify the real and fake news for covid-19 depends on the polarity of the corresponding data set. Finally, we achieved 94.49 % accuracy with the Random forest classifier. Also, we achieved 92.07% accuracy with the decision tree classifier; all of its results were great with our model.

**Keywords:** Coronavirus covid-19, Random Forest, Text Classification, Fake News, Natural Language Processing, binary classification, Machine Learning Algorithms, Decision tree, text mining.

## 1. INTRODUCTION

World Health Organization (WHO), in December 2019, advertised about Pneumonia unknown cause in a city of Wuhan, China. Those cases did diagnose as having acute Pneumonia and dryness, tiredness, fever, cough, and breathing difficulty. In January 2020, WHO called the virus "Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2)" [6]; after that called Coronavirus COVID-19, this is a virus spread all over the world to become a global epidemic. The widespread use of social media has become there misleading the public with Fake news. Fake news means untrue facts displayed, such as news, which threatens the reliability concerning inscribing, presented, and published by using social networks as true news. Also, during that digital era we live in now, social media is the easy way using and tremendous hurdles in order to restrain and control the spread of misleading for a number of reasons; one of this free flow of data and information into different positions social media such as Twitter, YouTube, Facebook, Instagram, and all. During the situation, an article report tells, approximately one in two citizens crossed the United Kingdom U.K., United States US, Spain, Argentina, Canada, South Korea, and Germany, claim they have seen mislead information on the social networks associated with coronavirus virus(covid-19). Because of a teeming spread of fake and misleading news, some commentators hold immediately referring to the latest

wave of misinformation accompanied by the covid-19 pandemic [7]. As knowledge becomes the approach, we understand the world. Hence, false knowledge and misleading news can use on ideas. A study in intelligence confirms that the falsehoods opening in social networks have an especially strong impact on less realizes people also prevent them from obtaining the right arrangements [8]. Fake news is utilized to make people racist and raise ideas, fears, head to bullying, including disturbing the general tranquility of society and brutality against innocent people. Despite false news not only in COVID19 even but becomes a big democratic impact [23]. The fake news was happening a big influence in the U.S. presidential election, how prejudiced conviction and people's ideas [9]. Recently become many disturbances and unfortunate incidents in Arab countries and other countries of the world destroy innocent people, seven murders caused by the communication sites because of fake news [10]. Because of the increasing and increasing numbers of patients into web site based active, data news could be quickly shared in any user, away from reliability. Since fake news is written to sow trick the readers, it gives it a hard duty to identify detect based on the content data. Algorithms machine learning detection necessary to get an expert method to identify detect misleading as news contents to change into names of the way that is composed. Fake news detects one of the applications of the natural language process NLP has recently received significant attention from specialists and researchers. So greatly, several methods have been involved in the detection of fake news [11]. Most maximum of a method proposed into the article paper to recognize fake news understanding by intricacy or problems because of a classify assignment with linking labels such as false or true or fake or real, etc. with a special text. In most cases, machine learning ML techniques are used to achieve hopeful results. Analysis proves the SVMs have outperformed a number of machine learning supervised processes for deception detected into a text by utilizing content-based extracted features [12]. However, a whole effort has been made. In this article, the detection of misleading or fake news is offered used the use of machine learning algorithms; content-based features have been extracted from the binary text data set Coronavirus (COVID -19) download from figshare web set to classify the fake news [15]. The rest of that article paper is constructed as follows. In part, II presented a Literature Review of the related works. Part III discusses the methodology and data set analysis, i.e., whereby text data are gathered and preprocessed, feature extracted approach, including detect technique. Part IV describes the result achieved within our research then lastly, Part V represents the conclusion by viable future research orientations.

## 2. Literature Review

Notwithstanding the fake news and misleading detection issue that have been considered during the first time about Coronavirus so lately, that have dragged and attracted unusual observation. Different methods have been introduced in order to detect fake news COVID-19 into different kinds from unstructured data from social media and other sources. Related works and reviews of enduring in the discussion concerning fake news detect possess shown in this part. Dadgar et al. [1] the process proposed to classification news within different categories used Term Frequency-Inverse Document Frequency (TF-IDF), Support Vector Machine (SVM). Ozbay et al.[2] In their work, preliminary evaluation of the twenty-three brilliant analysis methods has been implemented in existed known datasets and those classifier model has been associated depended onto four methods evaluations metrics. Pirouz et al. [3] suggested a binary classification design presents a higher performance capability to predict the approved cases. Then, a regression study method has been produced, moreover that aim of approved cases and associated with variations from regular weather parameters (humidity, average, and wind temperature). The outcome result showed the related humidity; also, the highest regular clod temperature had the most extensive impact on the sanctioned events. The relative humidity in the foremost

case study, on average, 77.9%, changed absolutely. Reddy et al. [4] discussed methods for the detection of fake news used just feature extraction about the text of news without used different associated metadata. Noticed a compound of stylometric features extraction also text word vector illustrations within ensemble techniques. Samuel et al. [5] have identified common sentiment attached within the pandemic used Coronavirus (COVID-19) Tweet and R analytical software, within its sentiment investigation insight in this progress from fear-sentiment extra time as COVID-19 approach top level into the USA United States, used clear textual analytics carried through needed text data visualization, they did two fundamental ML classification techniques, within the meaning of text analytics, moreover that compared their efficiency during classified, they found 91% for short Tweet, including the Naïve, and accuracy of 74% with a shorter tweet.

### **3. Fake news detection model**

This part presents the facts of the offered models during fake news covid-19 detection. We start through preprocess the data set through cleaning the unnecessary terms call for records like stop-words, numbers, etc. Next, Feature extraction becomes implemented to the Covid-19 fake news dataset for decreasing the dimension of feature time. The expression in all text documents is the Document-Term Model is created and weighted. The latest technique of a model is to implement an artificial intelligence algorithm supervised at Covid-19 fake news dataset [13].

#### **3.1. Text Data mining**

During the learning method, a summary and illustration of text data considerably affect this accuracy of outcomes. Particularly, the text sentiment analysis challenges require to obtain conversation within the description that is fitting to the system to be utilized. Text mining -based text received by the users at a social network are usually into unstructured data. For this purpose, extracted structured data from the social networks must be converted within an unstructured form text data mining process. The difficulty of text data can be described as an information mining of helpful, significant, including earlier private knowledge of the test text dataset [14, 21].

## **4. PROPOSED METHODOLOGY**

This part provides a schematic diagram for this proposed fake news detection. It starts through gather, data preprocessing the text data set next cleaning text data through removing the unnecessary phrases, numbers like stop-words, characters, and etc. next one feature extraction models have been applied through several steps, the most important processed data in order to detect the fake news from text data set, in this step normalization convert from human language to machine language for a deal with algorithms decreasing the dimension of the feature term. The words per document are the Document-Term Matrix is built and weighted. The current method of a text model is to use artificial intelligence (A.I.) supervised machine-learning algorithm onto the COVID-19 fake news text dataset [17, 18]. Finally, we test and train our model to classification algorithms, specifically; Random forest classifiers and Decision tree classifiers, have been used in that research. Fig. 1 describes the method of the design model.

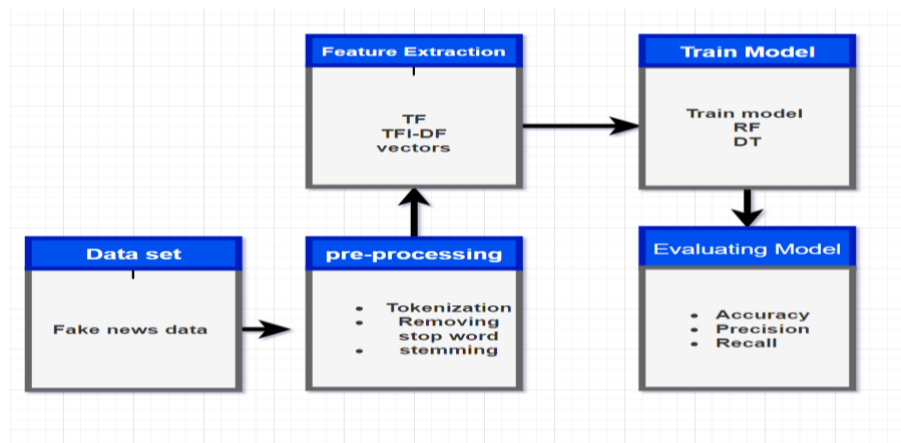


Figure 1. The process of the proposed model

#### 4.1. Data set Collection

During this study research, we have been applied to classify fake news of a Covid-19 binary classification data set. We data set to download from the web concerned with fake news from the Covid-19 binary classification data set; it is illustrated in Figure (2), and (3) shows Text and Label provide below [24].

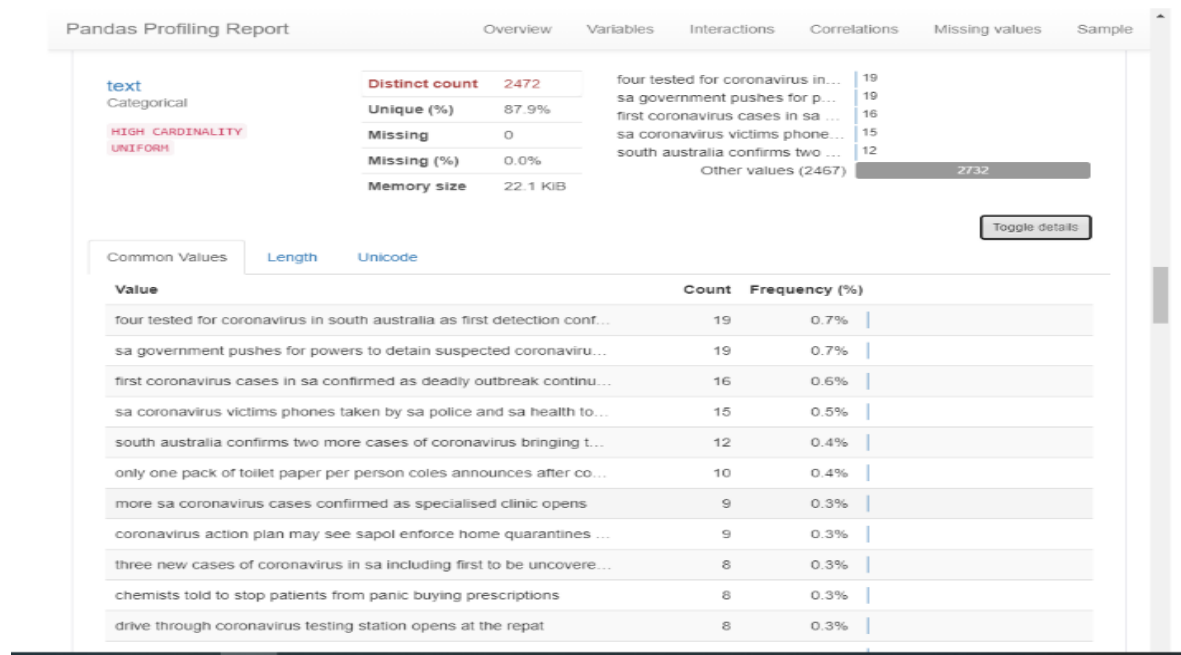


Figure 2. Text data set

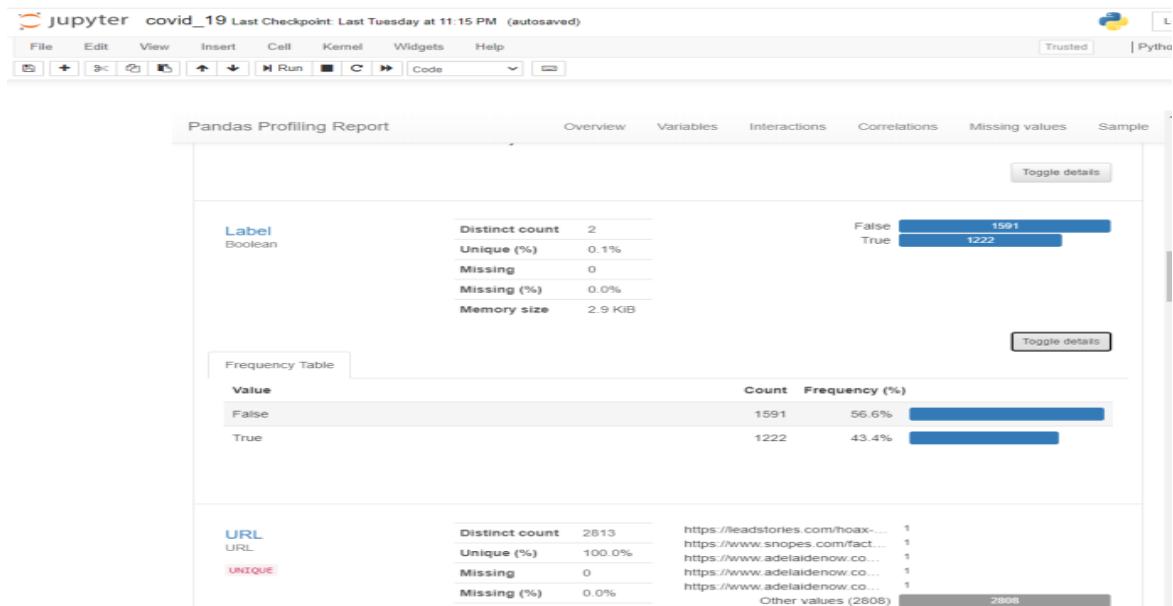


Figure 3. Label data set

## 4.2. Data Preprocessing

Preprocess is an important step in showing accurate outcomes. Preprocess the second step after collecting the data (raw text) dataset. Raw text data might contains symbols, characters, unnecessary special, and different things not necessary for our classification work [15].

Here might be some symbols suitable for sentiment analysis but for our work not suitable. We further raised different unique styles, e.g., @,!, #, Etc., of our text data. Those components can reduce to improve a performance model algorithms of a classifier and fixed for the classifier preprocess steps shown below.

---

Input: Text data set

Output: Preprocess text data set

---

1. Delete special characters from text.
  2. Remove numbers, punctuation marks from text data.
  3. Converter from capital letter.
  4. Clean symbols, numerical values.
  5. Applied state converter to text.
  6. Remove stop word.
- 

## 4.3. Feature Extraction

The feature extracting the fitting has a perfect impact on our model's performance of the machine learning classifier algorithms. Hence, we used in our work some feature extracting such as TF-IDF, Count Vectorizer, and Bag of a word dealing with text data. TF-IDF Vectorizer (term frequency-inverse document frequency), and Count Vectorizer are utilized to extract features from text data or data set that are used for classifiers before feeding that within an algorithm classification [20]. Then, Count Vectorizer generated vectors that have different dimensions because of the particular words of the corpus. Per

individual word becomes special dimension plus contains 0 in the specific dimension, including 1 during others, which really puts a frequency of per word. TF-IDF vectorizer features numerical descriptions for words, whether it there or not. Word is the weight with frequency, produced with the inverse document frequency of them. During single names, a word in order to show a great product, however, throughout must be given each weighting or significance barely in data set Coronavirus (COVID-19) words. That would point to enhanced production of classification [18].

#### 4.4. Training and Testing

Following extracted each vector's feature, next, we used that feature vector, including a label, in order to train our model. The trained classifier then determines that the label is associated with the data set of new situations within the feature of vectors. Following, the achievement concerning the proposed models is measured. During this analysis, Then we applied two various machine learning algorithms ML, specifically, random forest and decision tree [22].

#### 4.5. Classification

The dataset obtained divided into two sets, a training set, also a test set to fill in the algorithm classifier. The training set contains an 80% dataset, and the test set includes a 20% dataset. Between several classifiers, we applied two widely utilized Random Forest Classifier also Decision Tree Classifier by a long kernel into the analysis. The Random Forest Classifier there is a set of advantages to using the Random Forest Algorithm, but one of the central advantages is that it decreases the risk of overfitting and the needed training time. Farther, it offers a high level of accuracy. Random Forest algorithm works efficiently in a huge dataset and provides highly accurate predictions by calculating missing text data [25]. Random Forest a probabilistic classifier that is fast to implement. Random Forest is often used in a classification problem where the various occurrences of the word mean a set. On the other hand, The Decision Tree Classifier relates to the family of supervised learning algorithms that would be utilized for classification and regression difficulties [19]. It can handle both numerical data and categorical, Resistant to outliers; therefore, New features can be easily added, require little data preprocessing, and it can be used to build larger classifiers by using ensemble approaches [17, 34].

#### 4.6. Random Forest Algorithm 1

Random Forest. Algorithm two Random Forest obtains an organization learning process for regression, classification, also different tasks like a mixture of several decision trees; everyone has a unique node that used different text data in order to manage on modified. It connects the determinations of many decision trees to obtain the most desirable solution, which means a medium decision tree [33]. The random decision is a resilient, simple to train machine learning ML algorithm so as to usually make, indeed out, setting tuning, and an enhanced, and improved outcome result. That can be utilized concerning whoever classification and regression issue [30].

##### Input :

Text data set =Covid-19 Fake News text file

X:

x:

Y:

attributes \_records:

test attributes:

```

T
Output:
Vectorizer feature
Task Choose of feature (Text dataset)
  a. For i =0 to Y do
  b.   change easy _set = z
  c. Build node Y
  d.   Note tree(Y)
  e. End for
  f. Build tree(y)
  g.   IF Y = attribute record then
  h.     Return(Y)
  i. Else
  j.   Choose from Vectorizer feature
  k.     Choose Vectorizer F.
  l. For l = ( 0) to t do
  m.   Estimate taster Y to X, where X is feature = correspond Vectorizer named tree (Y)
  n.   End for
  o.   End if
End Task

```

**Algorithm 1: Random Forest.**

#### 4.7. Decision Tree Algorithm 2

Decision Tree. Algorithm number one is applied in order to split a large collection from text data documents into constant subsets like records through using a series of easy decision rules. Algorithm one divides the highlighted feature area into subset wherever per subset consists of a similar collection. This result means the tree, including decision node and leaf node highest decision nodes, into a tree that communicates on a most reliable predictor, which is named a root node.

The decision tree capacity handles all numerical data, variables, and categorical [32]. Here the decision tree used the knowledge accumulation theory through choose the best partitioning characteristic or attribute from that dataset. Know (Y) is measured used [31]:

$$\text{Know (Y)} = \sum_{i=1}^n P(y) \log_2 p(y)$$

A key benefit concerning the decision tree is it's simply executed. Then Decision trees Also, it's easy to interpret and understand the basic principle they're working on compared to other machine learning algorithms.

```

Input text data:
  Train text data = sequence of data calls
  X:
  x:
  long:
  attribute_ record:
  test _attribute:
Output:
  Vectorizer feature

```

### Task

- ```

a.   Build a node N
b.   If N = x Then
c.       return(m)
d.   Else
e.       X = m
f.   End if
g.   If attribute _record = 0 Then
h.       return(m)
i.   Else
j.       X = long
k.   End if
l.   test _attribute = long
m.   For ai To test _attribute
n.   14. Simplest = N is section.
o.   If test _attribute = ai
p.   m = test conditional
q.   End if
r.   End For Loop

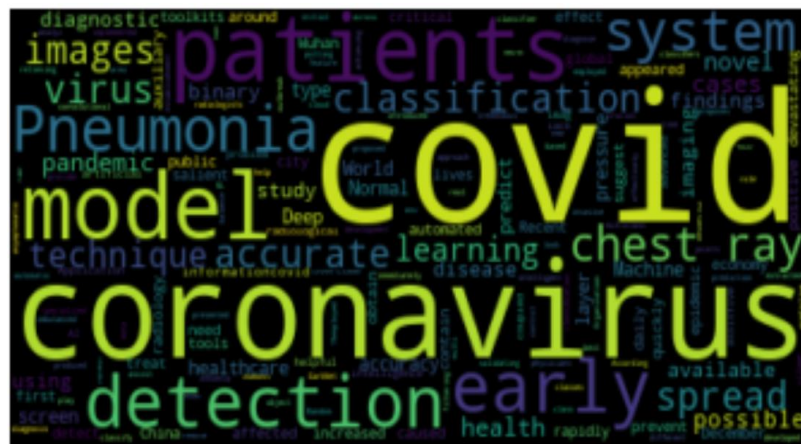
```

## End Task

**Algorithm 2: Decision Tree.**

## 4.8. Word cloud

Several times you might have seen a cloud loaded with many words in various sizes, which the importance of each word or represent the frequency. This is called Word Cloud or Tag Cloud. In the figure shown below, you will see how many words related to the COVID-19 data set, as shown in the report more lively [16].



### Figure 4: COVID-19 data set

#### 4.9. Performance evaluation metric

Several evaluation metrics possess used to associate the showing and performance of artificial intelligence supervised algorithms concerning covid-19 fake news detection. Then the evaluation metric is frequently utilized within an artificial intelligence supervised algorithm also authorizes us to test the efficiency of model algorithms. Here a



confusion matrix becomes applied to estimate achievement covid-19 fake news detection is displayed within Table 1. Within this matrix, an example is labeled because of real fake [26, 27].

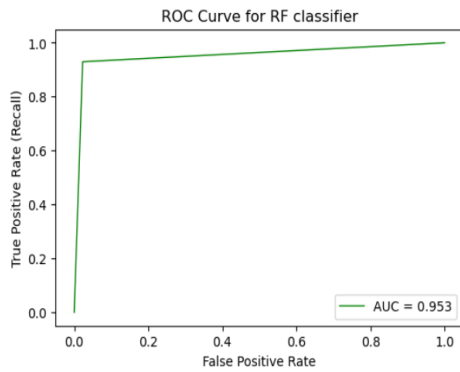
**Table 1. The evaluated metric for the Covid -19 detection method.**

| # | Evaluation   | Formula                                                       | Definition                                                                                                              |
|---|--------------|---------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------|
| 1 | Accuracy ACC | $\frac{TP + TN}{dxTP + TN + FP + FN}$                         | Is the percentage of accurately predict news on all of the examples.                                                    |
| 2 | Precision P  | $\frac{TP}{TP + FP}$                                          | Measures metric the covid-19 fake news then is accurately foretold of the total number predicted news in a fake label.  |
| 3 | Recall R     | $\frac{TP}{TP + FN}$                                          | Confirms the valuation of fake news, which precisely predicts the total quantity of fake news.                          |
| 4 | F1-Score     | $\frac{2 \times PRECISION \times RECALL}{PRECISION + RECALL}$ | F1-score is a harmonious medium value of precision value and recall value achieved during covid-19 fake news detection. |

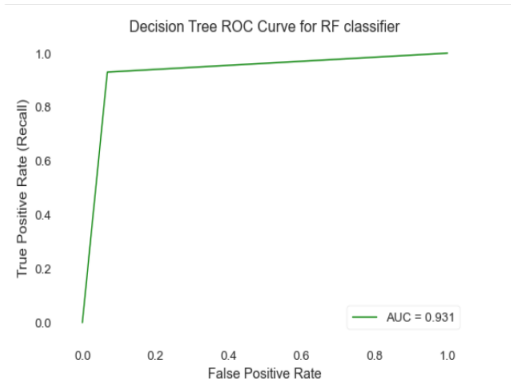
During the fake news detection difficulty, a first accuracy percentage of accurately predict news on each of the examples. The second Precision measures metric the covid-19 fake news then is accurately foretold of the total number predicted news in a fake label. The third recall rate confirms the valuation of fake news, which precisely predicts the total quantity of fake news. Finally, F1-score is a harmonious medium value of the precision value and recall value achieved during covid-19 fake news detection [28].

## 5. RESULTS AND DISCUSSION

During this stage, the experimental and implementation code for our modified model algorithms, we improve results as discussed and shown below. We collected data text datasets then analyzed them to understand our text data. Next, we test our model by many steps start with preprocessing to test model passing through feature engineering T.F.- IDF and Bag of a word, and classification procedures implied implemented to classify this detection of Covid-19 Fake News text data. Lastly, the production performance evaluation tests and measures obtained applied to evaluate the representation of the classifier. The performance analysis of Random Forest and Decision tree classifiers on our text dataset has been defined. The dataset comprised split into two sections randomly. The first section contains 80% of the data, which is utilized for training a classifier, and the remained 20% is to test. Random Forest gave us 94.49% accuracy for our dataset, whereas the Decision tree Classifier gave 92.07% accuracy. We noticed that Random Forest's performance better than the Decision tree in overfitting, as shown in figure 5, 6 below.

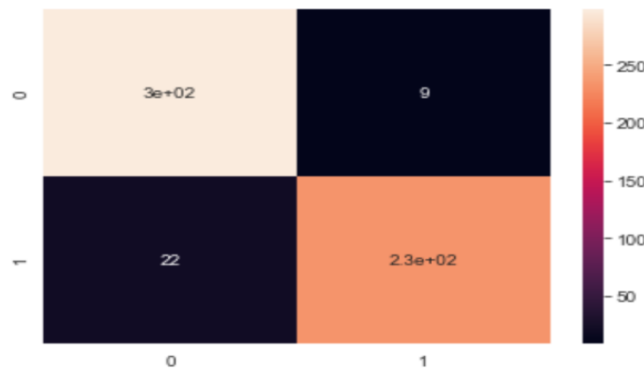


**Figure 5. Random Forest ROC Curve**



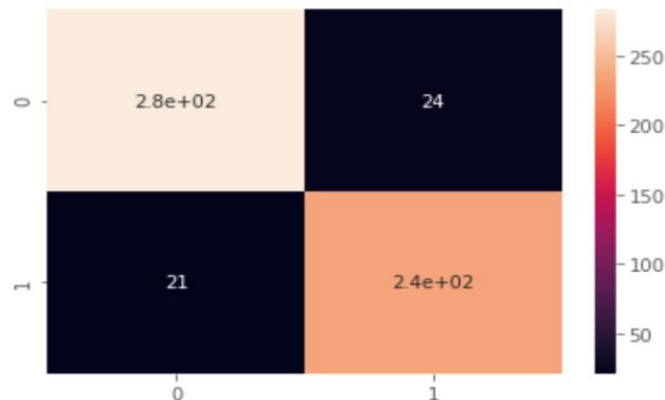
**Figure 6. Decision tree ROC Curve**

Figure 5 presents the confusion matrix for Random Forest that shows 298 news was shown, such as real and was really real news. On the other hand, 23 news was prophesied as fake, which acts truly real, which means false negative. Likewise, 233 fake news obtained predicted as fake also was truly fake 9 nine occurred predicted.



**Figure 7. Random forest Confusion Matrix**

Figure 6 presents the confusion matrix for Decision tree classifiers that explicates at 283 news occurred registered, such as real and was actually real news. Any other way, only 21 news was predicted as fake, which happened truly real, which means false negative. Then we can see the Decision tree produces less false negative. Furthermore, 235 fake news was predicted as fake and occurred truly fake just 24 did predict as fake and obtained truly real, which newly obtained fake shown such as real.



**Figure 8. Decision tree Confusion Matrix**

We are experimental that Random forest performance better than the Decision tree classifier. Table [2] and [3] display the F1- score, precision, recall, and support to Random forest then Decision tree classifier models sequentially.

**Table 2. Random Forest Classifier Results.**

|       | precision | recall | f1-score | support |
|-------|-----------|--------|----------|---------|
| False | 0.93      | 0.98   | 0.96     | 307     |
| True  | 0.97      | 0.92   | 0.94     | 256     |
| Avgas | 0.95      | 0.95   | 0.95     | 563     |

**Table 3. Decision tree Classifier Results.**

|       | precision | recall | f1-score | support |
|-------|-----------|--------|----------|---------|
| False | 0.94      | 0.93   | 0.94     | 307     |
| True  | 0.92      | 0.93   | 0.92     | 256     |
| Avgas | 0.93      | 0.93   | 0.93     | 563     |

## 6. CONCLUSION

In this research article, we proposed to detect Covid-19 Fake News from the social networks by means of combining text data mining techniques used supervised machine learning ML algorithms. Text mining analysis and classification process have been handled that text data set. During the experiment, two various algorithms, specifically, random forest and decision tree, are utilized to detect Covid-19 Fake News and modified model decision tree, between two algorithms, executed achieved greatly in the expression of accuracy, precision, recall, f1-measure, and support. Our experimental test outcomes illustrate that the offered detect Covid-19 Fake News from social media classification's trained and tested accuracy has achieved at 94.49 % in Random Forest and, in a decision, the tree has achieved at 92.07%. Researchers declared some facts about Fake News about community digital from during the rumors negative and liar. Finally, accord to the achieved outcomes, the greatest uses in terms of accuracy, precision, recall, and support have been collected from Random Forest Classifier algorithms. The decision tree classifier has also given closer results to Random Forest Classifier.

## REFERENCES

- [1] S. M. H. Dadgar, M. S. Araghi, and M. M. Farahani, "A novel text mining approach based on TF-IDF and support vector machine for news classification," *Proc. 2nd IEEE Int. Conf. Eng. Technol. ICETECH 2016*, no. March, pp. 112–116, 2016, doi: 10.1109/ICETECH.2016.7569223.
- [2] F. A. Ozbay and B. Alatas, "Fake news detection within online social media using supervised artificial intelligence algorithms," *Phys. A Stat. Mech. its Appl.*, vol. 540, p. 123174, 2020, doi: 10.1016/j.physa.2019.123174.
- [3] B. Pirouz, S. S. Haghshenas, S. S. Haghshenas, and P. Piro, "Investigating a serious challenge in the sustainable development process: Analysis of confirmed cases of COVID-19 (new type of Coronavirus) through a binary classification using artificial intelligence and regression analysis," *Sustain. (United States)*, vol. 12, no. 6, 2020, doi: 10.3390/su12062427.

- [4] H. Reddy, N. Raj, M. Gala, and A. Basava, "Text-mining-based Fake News Detection Using Ensemble Methods," *Int. J. Autom. Comput.*, vol. 17, no. 2, pp. 210–221, 2020, doi: 10.1007/s11633-019-1216-5.
- [5] J. Samuel, G. G. M. N. Ali, M. M. Rahman, E. Esawi, and Y. Samuel, "COVID-19 Public Sentiment Insights and Machine Learning for Tweets Classification," *Information*, vol. 11, no. 6, p. 314, 2020, doi: 10.3390/info11060314.
- [6] S. Latif et al., "Leveraging Data Science to Combat COVID-19: A Comprehensive Review," *TechRxiv. Prepr.*, no. April, pp. 1–19, 2020, [Online]. Available: <https://doi.org/10.36227/techrxiv.12212516.v1>.
- [7] L. Waikhom and R. S. Goswami, "Fake News Detection Using Machine Learning," *SSRN Electron. J.*, vol. 4, no. 7, pp. 2019–2021, 2019, doi: 10.2139/ssrn.3462938.
- [8] I. Abaker et al., "A Machine Learning Solution Framework for Combatting COVID-19 in Smart Cities from Multiple Dimensions," *Medrxiv*, p. 2020.05.18.20105577, 2020, doi: 10.1101/2020.05.18.20105577.
- [9] R. Kumar et al., "Accurate Prediction of COVID-19 using Chest X-Ray Images through Deep Feature Learning model with SMOTE and Machine Learning Classifiers," *medRxiv*, p. 2020.04.13.20063461, 2020, doi: 10.1101/2020.04.13.20063461.
- [10] Wim Naudé, "Artificial Intelligence against COVID-19: An Early Review," *Toward. Data Sci.*, no. March, pp. 2019–2020, 2020, [Online]. Available: <https://towardsdatascience.com/artificial-intelligence-against-covid-19-an-early-review-92a8360edaba>.
- [11] T. Ozturk, M. Talo, E. A. Yildirim, U. B. Baloglu, O. Yildirim, and U. Rajendra Acharya, "Automated detection of COVID-19 cases using deep neural networks with X-ray images," *Comput. Biol. Med.*, vol. 121, no. April, 2020, doi: 10.1016/j.compbimed.2020.103792.
- [12] C. T. Mesquita, A. Oliveira, F. L. Seixas, and A. Paes, "Infodemia, Fake News and Medicine: Science and the Quest for Truth," *Int. J. Cardiovasc. Sci.*, pp. 0–2, 2020, doi: 10.36660/ijcs.20200073.
- [13] M. G. Hussain, M. R. Hasan, M. Rahman, J. Protim, and S. Al Hasan, "Detection of Bangla Fake News using MNB and SVM Classifier," 2020, [Online]. Available: <http://arxiv.org/abs/2005.14627>.
- [14] J. Chen and S. Yan, "Verbal aggression detection on Twitter comments: convolutional neural network for short-text sentiment analysis," *Neural Comput. Appl.*, vol. 0, 2018, doi: 10.1007/s00521-018-3442-0.
- [15] [https://figshare.com/articles/COVID19FN\\_Dataset\\_csv/12489293](https://figshare.com/articles/COVID19FN_Dataset_csv/12489293)
- [16] <https://www.geeksforgeeks.org/generating-word-cloud-python/>
- [17] Siddiqui, N. A. Khan, and M. A. Khan, "PMKBEA: A Process Model Using Knowledge Base Software Engineering Approach," pp. 5–7, 2011.
- [18] S. Gilda, "Notice of Violation of IEEE Publication Principles: Evaluating machine learning algorithms for fake news detection," 2017 IEEE 15th Student Conference on Research and Development (SCORED), Putrajaya, 2017, pp. 110–115, doi: 10.1109/SCORED.2017.8305411.
- [19] U. Saeed, M. Sarim, A. Usmani, A. Mukhtar, A. B. Shaikh, and S. K. Raffat, "Application of Machine learning Algorithms in Crime Classification and Classification Rule Mining," vol. 4, no. 3, pp. 106–114, 2015.

- [20] Rabindra Lamsal, "Coronavirus (COVID-19) Tweets Dataset", *IEEE Dataport*, 2020. [Online]. Available: <http://dx.doi.org/10.21227/781w-ef42>. Accessed: Jul. 14, 2020.
- [21] A. Ahmad, T. Siddiqui, and N. A. Khan, "A Detailed Phasewise Study on Software Metrics : A Systematic Literature Review," vol. 3, no. 3, pp. 1696–1705, 2018.
- [22] L. Waikhom and R. S. Goswami, "Fake News Detection Using Machine Learning," *SSRN Electron. J.*, vol. 4, no. 7, pp. 2019–2021, 2019, doi: 10.2139/ssrn.3462938.
- [23] H. Reddy, N. Raj, M. Gala, and A. Basava, "Text-mining-based Fake News Detection Using Ensemble Methods," *Int. J. Autom. Comput.* vol. 17, no. 2, pp. 210–221, 2020, doi: 10.1007/s11633-019-1216-5.
- [24] M. K. Elhadad, K. F. Li, and F. Gebali, "Detecting Misleading Information on COVID-19," *IEEE Access*, vol. 8, pp. 165201–165215, 2020, doi: 10.1109/access.2020.3022867.
- [25] J. Abrahamsson, "Loan Default Prediction using Supervised Machine Learning Algorithms," 2019.
- [26] X. Zhou and R. Zafarani, "A Survey of Fake News: Fundamental Theories, Detection Methods, and Opportunities," 2018, doi: 10.1145/3395046.
- [27] H. M and S. M.N, "A Review on Evaluation Metrics for Data Classification Evaluations," *Int. J. Data Min. Knowl. Manag. Process*, vol. 5, no. 2, pp. 01–11, 2015, doi: 10.5121/ijdkp.2015.5201.
- [28] M. Baykara and B. Sekin, "A novel approach to ransomware: Designing a safe zone system," *6th Int. Symp. Digit. Forensic Secur. ISDFS 2018 - Proceeding*, vol. 2018-Janua, pp. 1–5, 2018, doi: 10.1109/ISDFS.2018.8355317.
- [29] S. Panda, "Detecting Twitter's Impact on COVID-19 Pandemic in India," *Int. J. Innov. Technol. Explor. Eng.*, vol. 9, no. 8, pp. 815–819, 2020, doi: 10.35940/ijitee.h6718.069820.
- [30] C. C. Aggarwal, X. Kong, Q. Gu, J. Han, and P. S. Yu, "Active learning: A survey," *Data Classif. Algorithms Appl.*, pp. 571–605, 2014, doi: 10.1201/b17320.
- [31] V. Klyuev, "Fake News Filtering: Semantic Approaches," *2018 7th Int. Conf. Reliab. Infocom Technol. Optim. Trends Futur. Dir. ICRITO 2018*, pp. 9–15, 2018, doi: 10.1109/ICRITO.2018.8748506.
- [32] J. Wan, P. Zheng, H. Si, N. N. Xiong, W. Zhang, and A. V. Vasilakos, "An artificial intelligence driven multi-feature extraction scheme for big data detection," *IEEE Access*, vol. 7, pp. 80122–80132, 2019, doi: 10.1109/ACCESS.2019.2923583.
- [33] B. Kondoch, "covid-19 and the Role of the Security Council as Global Health Peacekeeper," vol. 23, no. June 2020, pp. 105–116, 2019, doi: 10.1163/18754112-02301006.
- [34] T. Siddiqui, A. Y. A. Amer and N. A. Khan, "Criminal Activity Detection in Social Network by Text Mining: Comprehensive Analysis," *2019 4th International Conference on Information Systems and Computer Networks (ISCON)*, Mathura, India, 2019, pp. 224–229, doi: 10.1109/ISCON47742.2019.9036157.