Behavior Based Machine Learning Approaches to Identify State-Sponsored Trolls on Twitter

[1]A. Sarvani,[2]Shaik Anjum Javeariya, [3]Pathipati Swathi Lakshmi, [4]Venugopal Boppana
[1]sarvani.anandarao@gmail.com , [2]shaikanjumjaveriya@gmail.com
[3]swathipathipati99@gmail.com , [4]srees.boppana@gmail.com
[1]Assistant Professor,[2,3]Student, [4]Associate professor
[1,2,3] Lakireddy Bali Reddy College of Engineering, [4]NRI institute of Technology

Article Info
Page Number: 228 – 236
Publication Issue:
Vol. 71 No. 3 (2022)

Abstract
Social media is a powerful weapon that youth is much interested in due to the wider availability of internet in the recent times. Due to its wider availability, there is a need and necessity to keep track of the activity or interests that youth is much interested upon based on the individual activity to ensure that he is behaving right, that pays a new way of research in this field. For this case we have considered Twitter, social media site that a larger user base to conduct our study where the user data is retrieved and scrutinized based on the preferred language, tweets and preferences by other users and their social behavior etc. are considered for data analysis to understand various kinds of texts in different languages. The proposed work is further enhanced with the use of machine learning algorithms to determine the persons emotions and to eliminate the fake statements and estimate the kind of tweets whether considered as good or bad. Based on the provided training and test data using different algorithms based on machine learning such as SVM, Naïve Bayes and LDA algorithms the personality is analyzed firstly, later the use of Random Forest and extra gradient boosting techniques by connecting tweety through an API can determine the truthfulness of the statement provided by the individual achieving a greater accuracy of 90 and 99.61 percent accordingly for the estimation of performance analysis.

Keyword— Neural Networks, Machine Learning, Training and testing data.

Article History
Article Received: *12 January 2022*
Revised: *25 February 2022*
Accepted: *20 April 2022*
Publication: *09 June 2022*

1.INTRODUCTION

Social Networking is one of the crucial tools that plays an important role in capturing individuals' attention who are widely addicted to use the internet. A few groups of researchers have analyzed and discovered that the many individuals who use social media indirectly learn about the varied personalities of different people that they come across in the

day to day lives based on the other people's day to day activity such as commenting, tweeting, mutual connections and the type of language they prefer based on the interest. Based on our initial assumption that state- sponsored trolls cannot completely hide their suspicious behavior, we have developed eight features of user behavior on Twitter and used it with four machine- learning classifiers to identify state- sponsored trolls, including the Decision Tree, Random Forest, Adaboost, and Gradient Boost. The findings show that we can create a broad classification that can detect political trolls based solely on their actions, regardless of the content they post or the organization they work for. Adaboost and Decision in terms of accuracy and F1 score Tree are the most efficient of all ML models. Here particularly we have analyzed the behavior and focus our research on main two points where the first one includes twitters social behavior and the individual language habits and lastly include each personality dimension to anticipate the individual personality. Personality traits and traits have virtually identical relationships. Some basic standards make it easy to predict. We've come to the conclusion that data can be turned into information using multiple sets of categories like network size, tweet density, profession and number of connections. We looked at the predictable nature of the traits generated by personality trait projections. We've gone over a lot of data that is unsuitable for the next step in the process, to improve the accuracy and efficiency of the forecast.

Finally, we used a parallel machine learning technique[2,3] to move forward with the implementation model and see how far we can infer personality traits from tweets. We researched and selected the best algorithm for better forecasting. We use data from Twitter[1] in our study. The dataset for this research was collected from the Twitter social media site. These files are preprocessed before being sent to feature extraction and feature selection. Finally, machine learning algorithms train the extracted features and precisely classify the output personality. The data set is being collected from the Twitter API. It allows us to use the features of Twitter without the help of the website interface. The Twitter API can be used to post tweets or send messages automatically. The API is also referred to as the set of URL documentation. Through this, we have a lot of data to predict people's personalities. Before proceeding to the feature selection and training stage, the data of my personality was preprocessed. We have used Open NLP to pre-process the data. To get started, we've split the last word of each sentence, including punctuation and aggregation of similar words, using tokenization. We then deleted URLs, symbols, names, spaces and lowercase letters. The process in continued using few detailed steps which include checking the connections between users and their interactions in the social network. Secondly, using the XGBoost machine learning approach to highlight the social network properties between different categories. Further finding out how the categories are related within. Lastly, we determine an efficient method for analyzing and focusing the individual personalities.

2. LITERATURE SURVEY

In recent years, governments and political organizations have increasingly deployed state-sponsored trolls to disseminate public opinion on social media platforms through disinformation operations. [4] This problem has a negative impact on the democratic process and encourages mistrust. Through political systems, it sows discord in society and accelerates political polarization. As a result, an automated approach is needed to identify

sponsored-troll accounts on social media and reduce their impact on the political process, as well as protect [5] citizens from manipulation of opinions. In the distribution of digital information, [6,7] Twitter and other social networking and microblogging services play an important role. Despite the ubiquity and usefulness of social media, there is several cases of misuse. When corrupted users discover ways to profit from it, for example by increasing or decreasing the user's credibility status. There is no automated method for determining which news or users are reliable. As a result, developing a system that can assess a social media user's trustworthiness has become crucial. Aside from assigning a user's credibility score caught my interest.

Twitter data has been used extensively in many texts mining projects. Twitter's tweets, unlike those on other social media platforms, are public and easily accessible. Twitter provides a wealth of information. Russian- based online trolls successfully affected social media emotions during the 2016 US presidential election. While natural language processing approaches, we discover that are beneficial, yet they are insufficient. They are the only ones who can successfully identify online trolls. [8] A filtering strategy is insufficient to solve the problem. are beneficial, yet they are insufficient. They are the only ones who can successfully identify online trolls. A filtering strategy is insufficient to solve the problem.

Troll Hunter is an automated reasoning method that we deployed during theCOVID-19 conference to look for trolls on Twitter. A pandemic will occur in the year[9] 2020. Troll Hunter-Evader employs a Test Time Evasion (TTE) technique in conjunction with a Markov chain-based system[10,11,12] to recycle trolling tweets. When recycled tweets were used, the Troll Hunter's ability to correctly identify trolling tweets was lowered by 40%. We offer a detailed assessment of the repercussions of applying adversarial machine learning in the COVID- 19 epidemic since it has the potential to be deadly. figuring out how to avoid Twitter trolls,

The online new growing suspicious persons, generally referred to as trolls, are one of the most significant producers of hate, fraud, and deception on the internet. Some agendas are using these nefarious persons to spread inflammatory tweets, fooling the public. The difficulty in locating such accounts stems from the fact that they conceal their identities on social media, making it much more difficult to identify them only based on their social media profiles. As a result, we present a text-based strategy for detecting online trolls similar to those discovered during the US presidential election in 2016. Textual characteristics that use theme information, as well as profile features that identify accounts based on their behavior, are the foundations of our system.

3.METHODOLOGY

3.1 Dataset Description

In this proposed method, all data is taken from Twitter to get an opinion and is obtained in the form of a tweet. Following the selection of the Twitter data set, tweets were cleared of emotions and unnecessary punctuation, and a database was created to store the data in its converted version. In this method, all updated tweets are in lowercase alphabet and are divided into different parts of the tweet in a specific area. Words are tokenized to provide input into the machine learning process and then encoded as integers or floating-point values which is known as vectorization or feature extraction. Classifier is used to predict the results

of the test data, once it has been adequately trained, and then we compare proposed work performance with the existing work.

3.2 Dataset Description & Personality Classification

The Twitter API is being utilized to gather information without needing to go to the internet interface. The Twitter API can be used to automate the posting and sending of tweets and messages to predict people's personalities. For Personality Classification we employ Myers-Briggs personality profiling to determine a user's personality type. The MBTI is the most widely used method for character categorization, and it has been utilised in a variety of fields to assess a person's personality. The character is classified by the MBTI into four categories such as Introversion(I) or Extroversion(E), Intuition(N) or Sensing(S), Thinking(T) or Feeling(F), Perceiving(P) or Judging(J) With different types of Personality Indicators that include Introversion(I) or Extroversion(E), Intuition(N) or Sensing(S), Thinking(T) or Feeling(F), Perceiving(P) or Judging(J). The figure 1,2 shows the overview and work flow of the proposed study.

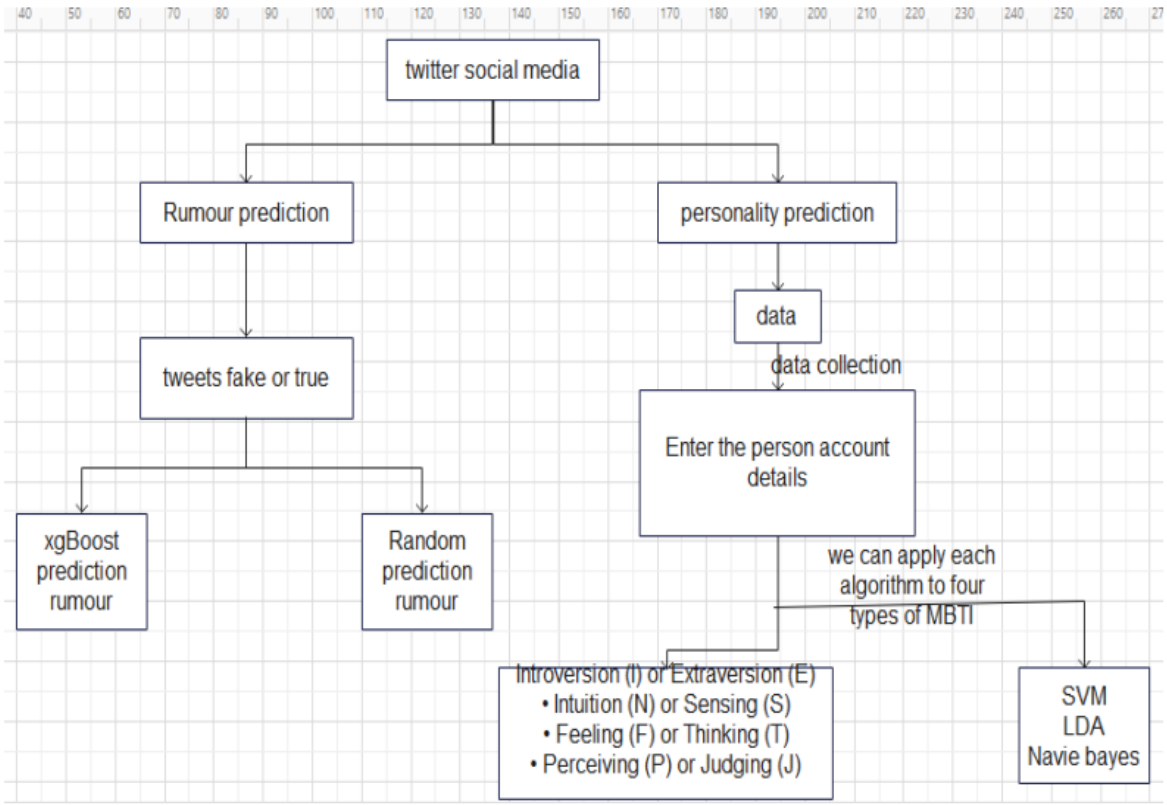


Figure 1: Overview of the proposed study

3.3 Proposed system steps

Step-1-Extraction of Data: To extract the viewpoint, data is first gathered and extracted from Twitter in the form of tweets. Following the selection of the twitter data set, the tweets were cleaned of emotions and unnecessary punctuation marks, and a database was created to store the data in its converted version. In this method, all of the updated tweets are in lowercase alphabet and divided into different chunks of tweets in the specific field. We'll go

over the specifics of the steps needed to alter data in the subsections that follow.

Step-2-Pre-processing of Data: Following are the Preprocessing steps that have been carried out are tokenization, removing punctuations and special symbols, stemming and lemmatization.

Step-3: Feature Extraction: People's behaviour on social media networks is influenced by the actions and presence of others on the platform. This could have an influence on the development of new data or activities via the groups. Many apps have been developed to better understand how these events occur and affect people. The dataset in ourstudy is divided into two categories: text feature extraction and image feature extraction. Text feature extraction is used to analyse people's language on Twitter. Topic and expression counts are two more aspects. CountVectorizer is a technique for gaining insight into people's psychology. It is commonly utilised in psychological studies. CountVectorizer was created to examine files in a number of formats Languages are translated in a timely and effective manner. TfidfVectorizer is a presently being tested analytic tool. It can be used to predict personality traits once it has fully grown. NLTK is used to extract features (Natural Language ToolKit). NLTK is a nlp library with packages for analysing human-machine communication. The package includes stemming, lemmatization, tokenization, POS tagging, and extraction.

Step-4: Predicting the personality of twitter account based on tweets Following feature extraction, train data is fitted to a suitable classifier, and once the classifier has been sufficiently trained, we use the classifier to predict the results of the test data, then compare the original value to the value produced by the classifier.

Step-5: Result Analysis

Our model, which is trained using the SVM[13,14] by classifying using the Myers Briggs Type Indicator, shows a greater accuracy of 90 percent of a person's personality. Compared with SVM(Support Vector Machine), LDA[15,16,17](Linear Discriminant Analysis), Naïve Bayes[18,19,20].

3.4 Work Flow

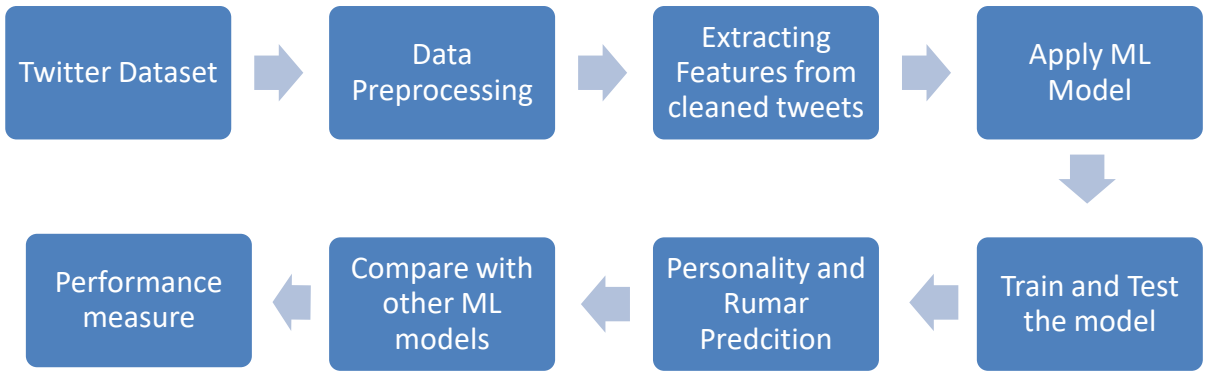


Figure 2: Work flow of the proposed study

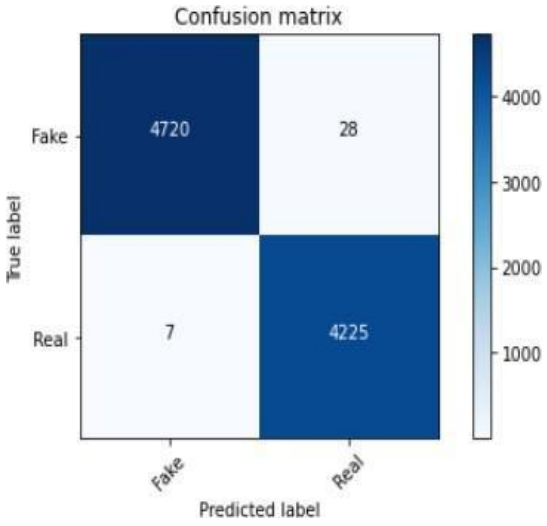
4. RUMOUR PREDICTION: We are predicting rumours that are fake or true by using two algorithms such as Random Forest, Extra Gradient Boosting. Our model which is trained using Extra Gradient Boosting shows a greater accuracy of 99% than Random Forest.

Step 1: Read Datasets.
Step 2: Data Cleaning and Data Preparation.
Step 3: Basic Data Exploration.
Step 4: Modelling.

EXTRA GRADIENT BOOSTING ALGORITHM: XGBoost is a scalable and highly accurate implementation of gradient boosting that pushes the limits of computing power for boosted tree algorithms, being built largely for energising machine learning model performance and computational speed.

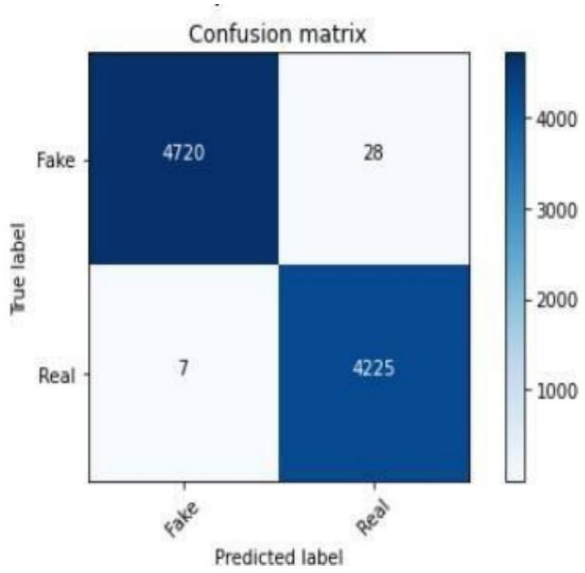
5.EXPERIMENTAL VALIDATION

Random forest and extra gradient boosting algorithms are used to anticipate rumors. We calculated the F1 score using multiple machine learning classifiers after training the model. After the extraction of the attribute, the train data is fitted for proper classification and once the classifier is adequately trained, the classification is used to predict the results of the test data. The figure 3 shows the confusion matrix by Extra Gradient Boosting Algorithm with accuracy of 99.6%. The figure 4 shows the confusion matrix by random forest with accuracy of 96.83%. Table 1 shows the accuracy between Extra Gradient Boosting Algorithm and random forest[21,22,23].



	predicted	percentage
fake	98	98.0
true	2	2.0

Fig.3 Confusion Matrix by Extra Gradient Boosting Algorithm



	predicted	percentage
fake	98	98.0
true	2	2.0

Fig.4 Confusion Matrix by Random forest

Table 1: Accuracy between Extra Gradient Boosting Algorithm and random forest

Extra Gradient Boosting Algorithm	Random forest
99.6%	96.83%.

Additionally SVM, LDA and Nave Baye s algorithms, are also used in calculation of personality predictions by using twitter data. The figure 5 shows the accuracy between the three algorithm, among these three SVM has given the maximum accuracy.

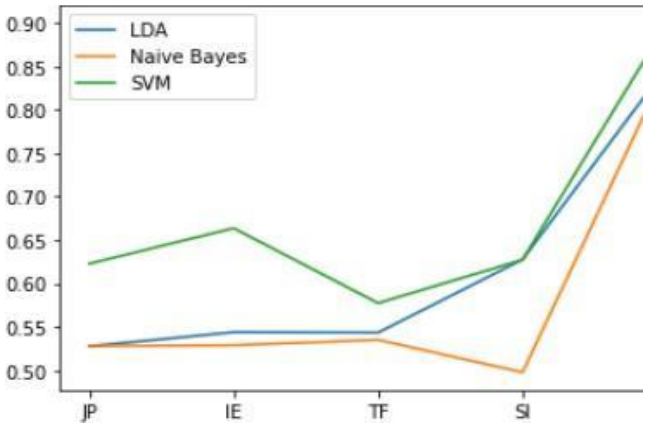


Fig.5 Accuracy Graph

The result clearly states that SVM has produced 90% accuracy in personality prediction and Extra Gradient Boosting has produced 99.61% accuracy in rumour prediction.

Using SVM, LDA, and Nave Bayes Algorithms, we calculated the personality prediction using Twitter data. We also used the Random Forest and Extra Gradient Boosting Algorithm to anticipate rumours. We calculated the F1 score using multiple machine learning classifiers after training the model. Following feature extraction, train data is fitted to a suitable classifier, and once the classifier has been adequately trained, the classifier is used to predict the results of the test data, and the original value is compared to the value produced by the classifier. The accuracy of different classifiers is shown below, and the best classifier with the highest accuracy percent is chosen. Several factors, such as the f1-score, mean, variance, and so on, are taken into account by the classifiers. We made a projection. Based on the performance study, the best conclusion was 90 percent accuracy using the SVM algorithm for personality prediction and 99.61 percent accuracy using Extra Gradient Boosting for rumour prediction.

6.CONCLUSION

This study has shown a framework of insights into each user's social network analysis and personality prediction. Using user perspectives on relationships, tweets , comments in social networks, we were able to conduct successful surveys on this routine study operation. Large number of datasets are used to determine a person's multiple personalities. Our findings show that examining the social and linguistic networks of personalities can yield a large number of insights. We have found a maximum of 90% accuracy in the SVM algorithm in personality prediction and 99.61% accuracy in predicting whether the tweet is fake or not using an extra gradient boosting algorithm.

REFERENCES

1. Anandarao, Sarvani, and Sweetlin Hemalatha Chellasamy. "Detection of Hot Topic in Tweets Using Modified Density Peak Clustering." Ingénierie des Systèmes d'Information 26.6 (2021).
2. Lavanya, K., et al. "Predicting The Emotions Based on Emoji’s and Speech Using Machine Learning Techniques." (2021).
3. Ray, Susmita. "A quick review of machine learning algorithms." 2019 International conference on machine learning, big data, cloud and parallel computing (COMITCon). IEEE, 2019.
4. Alhazbi, Saleh. "Behavior-Based Machine Learning Approaches to Identify State-Sponsored Trolls on Twitter." IEEE Access 8 (2020): 195132-195141.
5. Kannan Neten Dharan, Kannan Neten Dharan. "A comparative study of russian trolls using several machine learning models on twitter data." (2019).
6. Bilbao-Jayo, Aritz, and Aitor Almeida. "Improving Political Discourse Analysis on Twitter With Context Analysis." IEEE Access 9 (2021): 104846-104863.
7. Liu, Yinan, et al. "Named entity location prediction combining twitter and web." IEEE Transactions on Knowledge and Data Engineering 33.11 (2020): 3618-3633.
8. Saad, Shihab Elbagir, and Jing Yang. "Twitter sentiment analysis based on ordinal regression." IEEE Access 7 (2019): 163677-163685.
9. Sánchez-Corcuera, Rubén, Arkaitz Zubiaga, and Aitor Almeida. "Analyzing the Existence of

Organization Specific Languages on Twitter." IEEE Access 9 (2021): 111463-111471.
10. Das, Rahul Deb, and Ross S. Purves. "Exploring the potential of Twitter to understand traffic events and their locations in Greater Mumbai, India." IEEE Transactions on Intelligent Transportation Systems 21.12 (2019): 5213-5222.
11. Berk, M., et al. "Exploring the Potential of Twitter to Understand Traffic Events and Their Locations in Greater Mumbai, India.....".
12. Abdelminaam, Diaa Salama, et al. "ArabicDialects: An Efficient Framework for Arabic Dialects Opinion Mining on Twitter Using Optimized Deep Neural Networks." Ieee Access 9 (2021): 97079-97099.
13. Chauhan, Vinod Kumar, Kalpana Dahiya, and Anuj Sharma. "Problem formulations and solvers in linear SVM: a review." Artificial Intelligence Review 52.2 (2019): 803-855.
14. Mohammadi, Mokhtar, et al. "A comprehensive survey and taxonomy of the SVM-based intrusion detection systems." Journal of Network and Computer Applications 178 (2021): 102983.
15. Choubey, Dilip K., et al. "Comparative analysis of classification methods with PCA and LDA for diabetes." Current diabetes reviews 16.8 (2020): 833-850.
16. Maier, Daniel, et al. "Applying LDA topic modeling in communication research: Toward a valid and reliable methodology." Communication Methods and Measures 12.2-3 (2018): 93-118.
17. Ali, Liaqat, et al. "LDA–GA–SVM: improved hepatocellular carcinoma prediction through dimensionality reduction and genetically optimized support vector machine." Neural Computing and Applications 33.7 (2021): 2783-2792.
18. Chen, Shenglei, et al. "A novel selective naïve Bayes algorithm." Knowledge-Based Systems 192 (2020): 105361.
19. Berrar, Daniel. "Bayes' theorem and naive Bayes classifier." Encyclopedia of Bioinformatics and Computational Biology: ABC of Bioinformatics 403 (2018).
20. Chen, Shenglei, et al. "A novel selective naïve Bayes algorithm." Knowledge-Based Systems 192 (2020): 105361.
21. Speiser, Jaime Lynn, et al. "A comparison of random forest variable selection methods for classification prediction modeling." Expert systems with applications 134 (2019): 93-101.
22. Resende, Paulo Angelo Alves, and André Costa Drummond. "A survey of random forest based methods for intrusion detection systems." ACM Computing Surveys (CSUR) 51.3 (2018): 1-36.
23. Iwendi, Celestine, et al. "COVID-19 patient health prediction using boosted random forest algorithm." Frontiers in public health 8 (2020): 357.