

# Prediction of Terrorism and Threats Using Machine Learning

<sup>1</sup>K. Hemanthi, <sup>2</sup>D. Krishna Veni, <sup>3</sup>M. Anusha, <sup>4</sup>M. Nandini

<sup>1</sup>Assistant Professor, Department of Information Technology, Lakireddy Bali Reddy College of Engineering (Autonomous), Mylavaram, Krishna DT, Andhra Pradesh, India. hemanthikallam@gmail.com

<sup>2, 3 4</sup>B.Tech, Department of Information Technology, Lakireddy Bali Reddy College of Engineering (Autonomous), Mylavaram, Krishna DT, Andhra Pradesh, India.

## Article Info

Volume 83

Page Number: 2580 – 2588

Publication Issue:

May - June 2020

## Abstract:

The number of terror attacks is increasing globally from day to day and we have a need to analyze and predict the occurrence of the terror attacks. The effect of the terrorism increases mainly due to the internet, i.e., internet leaves as a platform to spread the terrorism in major 9/11 attacks in India, the attack created more havoc due to social media. So, when the government announces the major policies it will make the people to stay away from the social media to avoid terror attacks. We have a database called as GTD (Global Terrorism Database) which has the information about the terrorism activities. So, by using the information by this database, we can use some algorithms like Random Forest Algorithm, Gaussian Naive Bayes and Decision Tree Algorithm to predict and measure the accuracy of the occurrence of the terror attacks in may be future also. This will shows the list of countries that are involved in the major terror activities and the losses that are occurred to the countries till date due to terrorism.

**Keywords:** Global Terrorism Database, Data mining, classification algorithm's, processing techniques, python programming.

## Article History

Article Received: 11August 2019

Revised: 18November 2019

Accepted: 23January 2020

Publication: 10May2020

## I. INTRODUCTION

Terrorism is the word which is shaking the entire world. It is nothing but illegal use of the weapon, nuclear bombs and killing the large number of innocent people. Where are different reasons for which terror attacks are done like to stand globally as a supreme power in the world to conquer other places of the country for acquiring domination in the same community which occur among the different sets in the same community like Sunni's and Shia's to spread their religion all over the world. Not willing to live in the particular rule of a political leader for releasing their terrorist leader who was arrested in same other country. There are different types of terrorism like:

### A. State Sponsored Terrorism:

It is nothing but one state will sponsor money to their own terrorist in their country to attack another

state. There may be many reasons for the occurrence of this type of terrorism. The country that is mostly involved in this type of terrorism is Pakistan's former president "Musharraf" told that "there has been a training for their militant groups in their country to fight against India in Kashmir" this is one of the example of the State Sponsored Terrorist.

### B. Dissent Terrorism:

It is nothing but the international rebellion. There are some groups which didn't like the rule of their government. So, they will start violence like bombing attacks for killing the innocent people etc. This is one of the dangerous type of terrorism in order to make another person as a leader of the government, these rebellion groups will kill the other political leaders. Terrorism activities by particular group in order to gain special benefits from government, these two are the main reasons for the occurrence of Dissent terrorism.

### ***C. Terrorist left and right:***

There are some groups which are involved and rooted in particular political ideology and religious ideology. If any state government tries to disturb their ideology, they will do the attacks.

### ***D. Religious Terrorism:***

This is most common and widely seen cause for the occurrence of terror attacks in order to spread their religion across the world and to maintain their religion all over the world, attacks were done.

### ***E. Criminal Terrorism:***

This type of terrorism that is done by the criminals for the sake of making money. For e.g., Smugglers are involved in this type of terrorism. It also comes under internal terrorism. So, these are main causes and types of terrorism in the world globally.

One more interesting fact about terrorism is that, the amount of money that is invested to do terror attacks i.e., huge amount of money is required for performing the terror attacks. Because, it requires raw material gathering for making bombs planning for making illegal passports to enter into the other country. A lot of consequences will occur after the terror attack are happened like many innocent people were killed, huge economic loss, loss of tourism and a great loss of markets. So, we need to predict the occurrence of the terror attacks and for that process we use a machine learning algorithms to predict the terrorist attacks globally.

## **II. Literature Survey**

A lot of research had made with different types of problems involved in terrorism. So, came up with different machine learning algorithms. Accuracy of prediction is totally based on the input data which are in the form of data sets.

Terrorism [1], [2], [7] is pinning out each and every state in the entire world and killing the large number of innocent people. Where there are different reasons for which terror attacks are done like to stand globally as a supreme power in the world to conquer other places of the country for acquiring domination

in the same community which occur among the different sets in the same community like Sunni's and Shia's to spread their religion all over the world. Not willing to live in the particular rule of a political leader for releasing their terrorist leader who was arrested in same other country.

In light of expand of terror percentage throughout the years, we should deal with an enormous amount of terrorism information stored in warehouses which would be hard to be examined physically, and furthermore now a day's, terrorists are getting advanced in technology, so they utilizing advance techniques to keep government in front of them. The main focus is on the survey of algorithms and techniques utilized for identifying the terrorism.

Terrorism analysis [6], [8] is a methodology for recognizing where the more terrorism is taking place. The terror regions are for the most dependent on the crime type these distinguished terror zones are useful to diminish the terror percentage. Here in this paper the reason for gathering all this information is to get the knowledge actually what terrorism is, what are the types, in which areas the more terrorism takes place. So that it is easy to analyze and predict the states of terrorism by using different machine learning algorithms [3], [4], [9], [13], [19] like here in this paper the algorithms used are Decision Tree Algorithm, Gaussian Naive Bayes Algorithm and Random Forest Algorithm. The cloud technology [14], [15], [16], [17], [18] is used to store the large amounts of data and process it.

Terrorism [5] data analysis can help to the Terror investigators to analyze the Terrorism. Based on classification and regression extract the Terrorism areas into a structured data. The reason for events of terrorism like terror details of person and other factors we are focusing mainly on terror factors of early years. K-NN order was utilized. In the present framework, Random Forest Algorithm, Gaussian Naïve Bayes algorithm and Decision Tree Algorithm will be used to analyze and predict the terrorism information.

Security has consistently been one of the most critical concerns. Government and security organizations are striving to prevent terrorism and protect their people. Challenge of managing huge measure of information has become a significant issue for all organizations. Consequently, a terrorism data system that can procedure huge amount of data in a brief time is required for analysts to know terrorism hotspots. Data pre-processing [10], [11], [12] is done in terrorism database and three methodologies for crime analysis and prediction is performed. These three methodologies are looked at, and results are affirmed with ground truth.

### III. Proposed System

#### Functional aspects of proposed system

The figure 1 shows the flow of work. It is divided into 3 sections:

1. Collection of Data.
2. Pre-processing of data.
3. Apply of Classification Algorithms.
4. Estimation of Result in Accuracy.

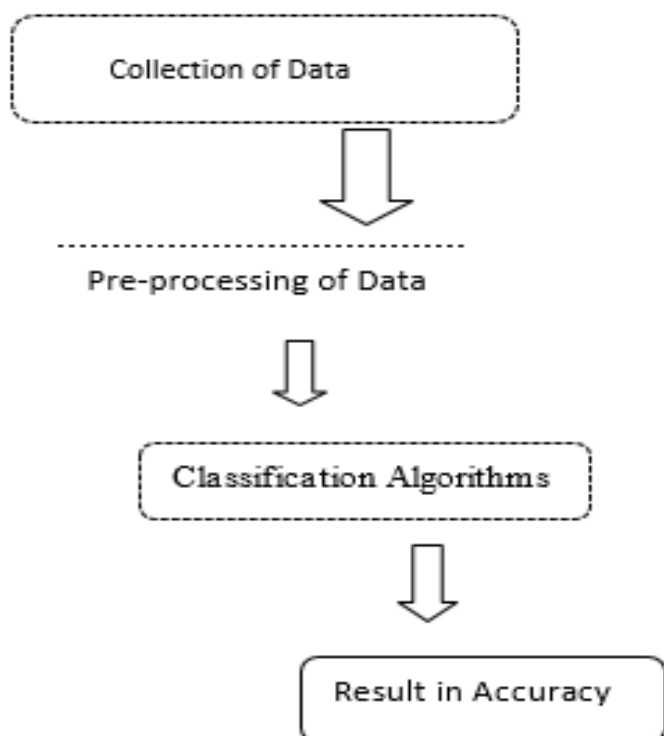


Fig. 1. Model Design.

#### A. Collecting Data

GTD is an organization where it will have all the data of the terrorist attacks which were happened in the past. They have the documents more than 190,000 international and domestic terrorism attacks which occurred worldwide since 1970. The data which is used to train and test the data is collected from the Global Terrorism Data Base. To starting of any process, it is very important to collect the data, here large datasets are collected for the better analyzation and prediction to find the better accuracy.

#### B. Data Pre-processing

The collected data from Global Terrorism Database will not follow a particular format, because, Terrorism can be in many forms: According to the National Advisory Committee, there are six types of terrorism. Where the entire attacks goal is to invoke fear and attempt to harm the lives of public. They are Politics Terrorism, Religious Terrorism, State terrorism, protesting terrorism, Quasi terrorism, Civil disorder terrorism. so, there will be a large data more than 170,000 instances collected through the investigation from the year started (1970) to till date. So, the data we can't directly use for training the data as it contains the missing values and loss of data. That's the reason we should do data pre-processing and then followed by training the data. Then all the missing values will be distracted and lossy data will be come out and the data which was driven from that Machine learning techniques is now ready for training the data.

#### C. Data Classification: Algorithms

##### Decision Tree Algorithm:

It is a type of data mining technique used to build classification models, regression problems and also used for categorical and numerical data types. Depending upon the training data sets we build training models which is used for prediction class labels. We compare the values of the root attribute with the recorded attributes. More class labels will

predict the more accuracy results, it is also used for removing over fitting problem.

### Random Forest Algorithm:

It is one of the classifier technique algorithm which is used in classification, regression problems, predictive accuracy and handling missing values. Here algorithm is not only used for training on different sets of data but also be used is different features to make decisions. In this algorithm the training data set is subdivided into training samples and those are used to build the decisions trees. More decision trees occurs more accuracy.

### Gaussian Naive Bayes' Algorithm:

The widening of the naive Bayes is called Gaussian Naive Bayes. it is also one of the Classification algorithm which is based on Bayes' theorem and one the simplest supervised approaches. it is used to calculate the probability of given information. The main advantage with this naive Bayes is it can also trained with small and large datasets.

### D. Estimation of Result in Accuracy

The results of the proposed method are discussed the following section.

## IV. Implementation and Results

In this paper we are using different classifier techniques, which are used for predicting accuracy. We finally went with random forest, decision tree and Gaussian naive Bayes. By grouping all the attacks information which is already happened and finding the accuracy upon them by using these three algorithms. By comparing the accuracy of these three algorithms random algorithm gives the best suitable algorithm.

### A. Visualization

In this visualization we are collecting different attributes from global terrorism database which is happened in the year 1996

to 2017. The following figure shows some of the attacks by various groups, deaths in years, countries and regions.

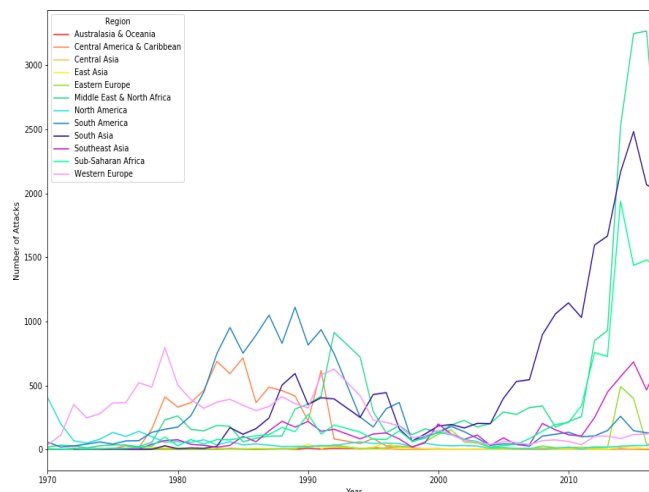


Fig. 2. Attacks by group name.

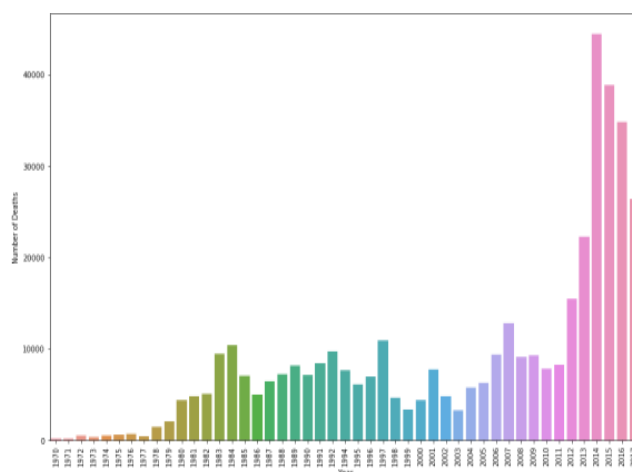


Fig. 3. Number of deaths by year.

The figure 2 shows no of attacks done by the group name called as Revolutionary armed force of Colombia (FARC) which is around 4500.

The figure 3 represents attacks happened from the year 1971 to 2017. Finally we conclude that attacks were continuously increasing from the year 2011 to 2014. From 2014 to 2017 continuously decreasing. In 2014 most of the deaths were occurred which is around 45000.

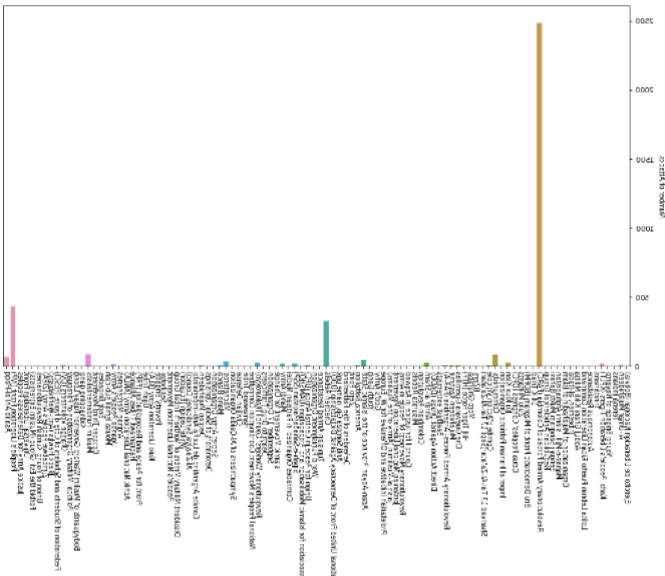


Fig. 4. Attacks in regions.

The figure 4 represents the number of attacks occurred in different regions. In this figure green color represents the region that is Middle East & North Africa which as most number of attackshappened.

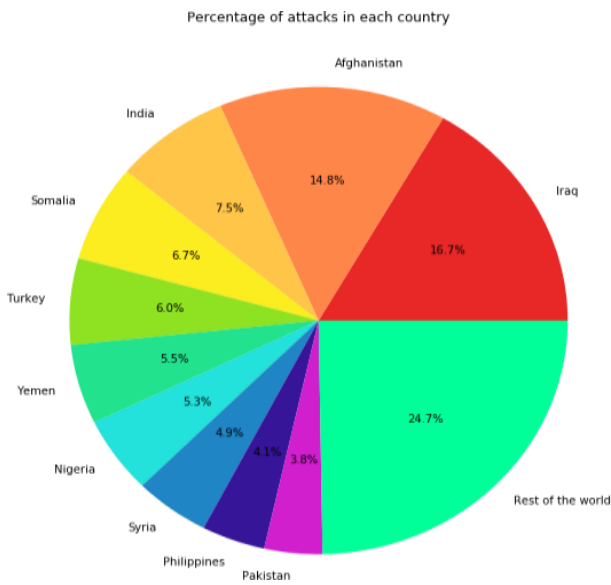


Fig. 5. Attacks in each countries.

## B. Classification and Regression

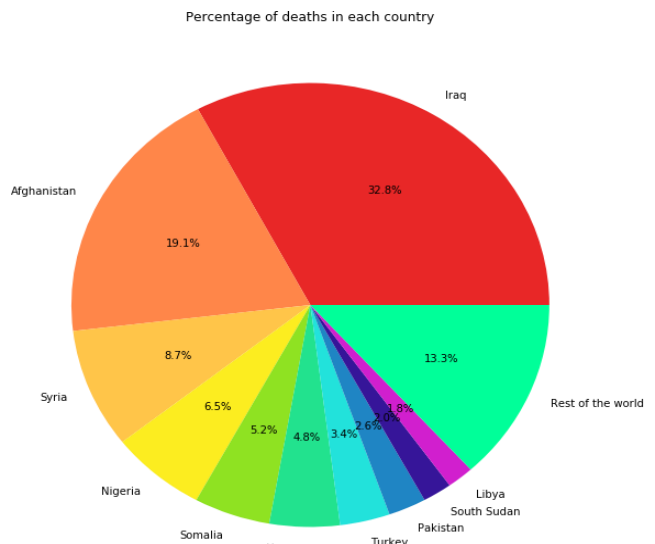


Fig. 6. Number of deaths in each country.

Some of the classification problems may get over fitting problem. To avoid this over fitting problem we have using different classifier techniques like random Forest, decision tree and Gaussian naive Bayes.

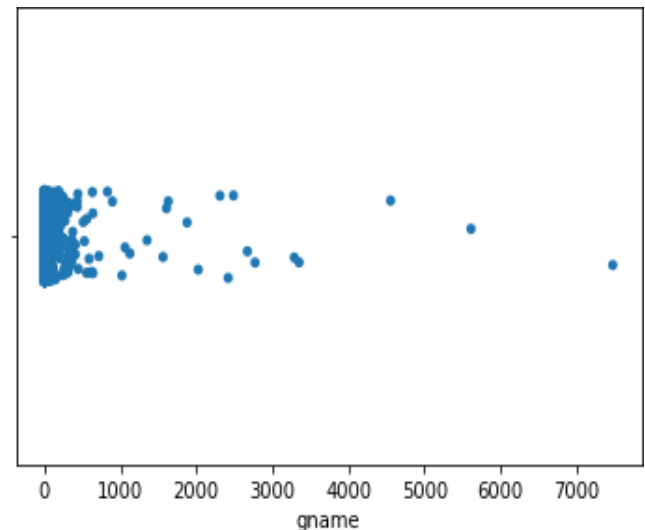


Fig. 7. Outliers problem.

In the figure 7 the region which is from 1000 to 7000 which leads a data will be over fitting problem. To avoid this problem we put some index point which is called threshold point. Lesser Threshold value will be result more accuracy value and lesser prediction values.



### C. Random forest Algorithm

In this random forest algorithm based upon the Threshold value we group the attributes of the terrorists which happened from the year 1970 to 2017 in figure 8. A list of more grouping values will be give more accuracy value.

```
[ ] res1=res.sort_values(by=['gname1'])
```

res1[250:300]		
35433	Abu Nidal Organization (ANO)	46
16375	Abu Nidal Organization (ANO)	46
16782	Abu Nidal Organization (ANO)	46
17398	Abu Nidal Organization (ANO)	46
19838	Abu Nidal Organization (ANO)	46
44241	Abu Nidal Organization (ANO)	46
13416	Abu Nidal Organization (ANO)	46
36708	Abu Nidal Organization (ANO)	46
33913	Abu Nidal Organization (ANO)	46
14505	Abu Nidal Organization (ANO)	46
68219	Abu Nidal Organization (ANO)	46

Fig. 8. Attacks by group name.

### C. Decision Tree Algorithm

Based upon the Threshold value we build the class labels. For example if the threshold value is 100, based upon the index value we create the decision trees which is class labels value equal to that of threshold. More class labels will be result the more accuracy values.

### D. Gaussian Naive Bayes

The widening of the naive Bayes is called Gaussian Naive Bayes. it is also one of the Classification algorithm which is based on Bayes' theorem and one the simplest supervised approaches. it is used to calculate the probability of given information. The main advantage with this naive Bayes is it can also trained with small and large datasets.

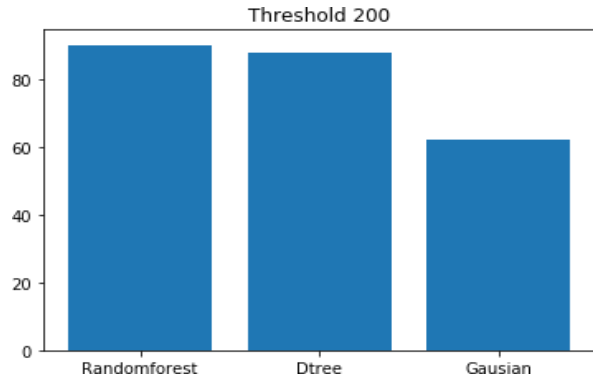


Fig. 9.a. Accuracy of algorithms.

The figure 9.a & 9.b shows the accuracy of different classification algorithms whereas threshold value is 200. In this representation random forest will result the highest accuracy value. The output of the above representation is as followsas:

```
[ ] rfscore=predict(rf,X_train, y_train, X_test, y_test)*1
print(rfscore)
```

```
88.17656332107198
```

```
[ ] dtscore=predict(tree,X_train, y_train,X_test, y_test)*
print(dtscore)
```

```
85.06568575932738
```

```
[ ] gnbscore=predict(gnb,X_train, y_train,X_test, y_test)*
print(gnbscore)
```

```
56.53179190751445
```

Fig. 9.b. Snapshot of the code for Accuracy of algorithms.

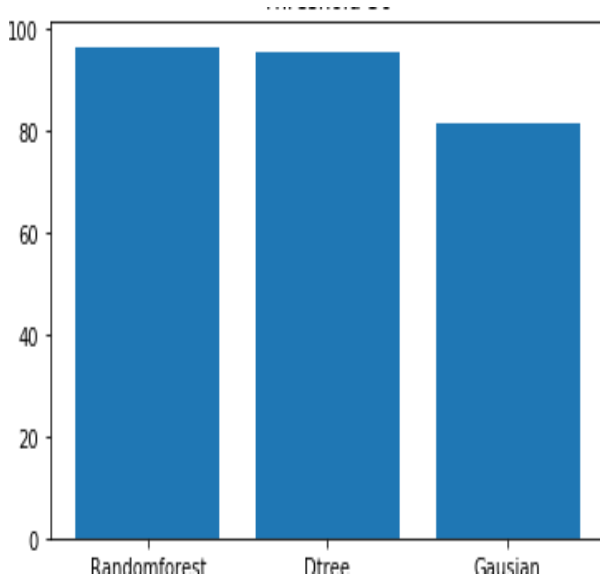


Fig. 10.a Accuracy of algorithms.

The figure 10.a & 10.b shows the accuracy of different classification algorithms whereas threshold value is 300. In this representation random forest will result the highest accuracy value. The output of the above representation is as follows:

```
[ ] rfscore=predict(rf,X_train, y_train, X_test, y_test)*100
print(rfscore)
```

96.27233518176217

```
[ ] dtscore=predict(tree,X_train, y_train,X_test, y_test)*100
print(dtscore)
```

95.20948860135552

```
[ ] gnbscore=predict(gnb,X_train, y_train,X_test, y_test)*100
print(gnbscore)
```

81.2230437461491

Fig. 10.b. Snapshot of the code for Accuracy of algorithms.

## V. Discussion and Conclusion

The dataset we used was from previous years analysis collected from the public database. The input data that is used as training datasets of Decision Tree, Random forest and Gaussian Naïve Bayes algorithm is plotted against multi-dimensional element which is portioned into different areas which are then categorized based on the order of training datasets. We collect the datasets from the Global Terrorism Database which is an organization where it helps for study for terrorism. After that we do a pre-processing of a data, it is a data mining technique that involves transforming raw data into understandable format. Data pre-processing is a method to resolve data incompleteness, data inconsistency, or lacking of trends and at last to the converted data we applying various classification algorithms because it helps you to make determination of an effective data classification process is important because it can help an organizations determine the appropriate level of control to maintain the confidentiality and integrity of data. It provides solid foundation for a data security strategy. A well planned data classification system makes essential data easy to find and retrieve. The goal of classification and regression problems is to accurately predict the target class for each case in the data. It belongs to the category of supervised learning targets also provided with the input data. By the above results it is clear that from the Random Forest Algorithm the accuracy is 96(approx.), from Decision Tree algorithm the accuracy is 95(approx.) and from Gaussian Naive Bayes the accuracy is 81(approx.). These outputs might change according to the Threshold Values.

## References:

- [1] Gary LaFree & Laura Dugan (2007) Introducing the Global Terrorism Database, *Terrorism and Political Violence*, 19:2, 181-204, DOI: 10.1080/09546550701246817.

- [2] T. Xia and Y. Gu, "Building Terrorist Knowledge Graph from Global Terrorism Database and Wikipedia," 2019 IEEE International Conference on Intelligence and Security Informatics (ISI), Shenzhen, China, 2019, pp. 194-196. doi: 10.1109/ISI.2019.8823450.
- [3] K. Indira, E. Brumancia, P. S. Kumar and S. P. T. Reddy, "Location prediction on Twitter using machine learning Techniques," 2019 3rd International Conference on Trends in Electronics and Informatics (ICOEI), Tirunelveli, India, 2019, pp. 700-703.
- [4] P. Karthika, R. Murugeswari and R. Manoranjithem, "Sentiment Analysis of Social Media Network Using Random Forest Algorithm," 2019 IEEE International Conference on Intelligent Techniques in Control, Optimization and Signal Processing (INCOS), Tamilnadu, India, 2019, pp. 1-5.
- [5] M. Luan, D. Sun and Z. Li, "Terrorism Risk Prediction Model Based on Support Vector Machine Optimized by Whale Algorithm," 2019 IEEE 7th International Conference on Computer Science and Network Technology (ICCSNT), Dalian, China, 2019, pp. 166-169.
- [6] K. Singh, A. S. Chaudhary and P. Kaur, "A Machine Learning Approach for Enhancing Defence Against Global Terrorism," 2019 Twelfth International Conference on Contemporary Computing (IC3), Noida, India, 2019, pp. 1-5.
- [7] H. Mo, X. Meng, J. Li and S. Zhao, "Terrorist event prediction based on revealing data," 2017 IEEE 2nd International Conference on Big Data Analysis (ICBDA), Beijing, 2017, pp. 239- 244.
- [8] M. Goel, N. Sharma and M. K. Gurve, "Analysis of Global Terrorism Dataset using Open Source Data Mining Tools," 2019 International Conference on Computing, Power and Communication Technologies (GUCON), NCR New Delhi, India, 2019, pp. 165-170.
- [9] A. Koneru, N. B. Naga Sai RajaniBhavani, K. Purushottama Rao, G. Sai Prakash, I. Pavan Kumar and V. Venkat Kumar, "Sentiment Analysis on Top Five Cloud Service Providers in the Market," 2018 2nd International Conference on Trends in Electronics and Informatics (ICOEI), Tirunelveli, 2018, pp. 293297.doi:10.1109/ICOEI.2018.8553970
- [10] K. Lavanya, L. S. S. Reddy and B. EswaraReddy , "A Study of High-Dimensional Data Imputation Using Additive LASSO Regression Model", Computational Intelligence in Data Mining, 2019 Advances in Intelligent Systems and Computing 711. (SCOPUS)
- [11] Lavanya K., Reddy, L., & Reddy, B. E. (2019). Distributed Based Serial Regression Multiple Imputation for High Dimensional Multivariate Data in Multicore Environment of Cloud. International Journal of Ambient Computing and Intelligence (IJACI), 10(2), 63-79. doi:10.4018/IJACI.2019040105.
- [12] Lavanya.K, L.S.S.Reddy, B. Eswara Reddy, "Multivariate Missing Data Handling with Iterative Bayesian Additive Lasso (IBAL) Multiple Imputation in Multicore Environment on Cloud", Volume 5, Issue 5 , May 2019, International Journal on Future Revolution in Computer Science & Communication Engineering (IJFRSCE), PP: 54 – 58.
- [13] Rama Devi Burri, Ram Burri, Ramesh Reddy Bojja, SrinivasaraoBuraga "Insurance claim Analysis using Machine learning Algorithms, "International journal of innovative technology and Exploring Engineering (IJITEE), Volume-8, Issue-6S4, April-2019, ISSN: 22278-3075.
- [14] A. Koneru and M. S. Latha, "Three tier architecture: To select CSP through CloudServiceBroker in multicloud environment," 2016 International Conference on Inventive Computation Technologies (ICICT), Coimbatore, 2016, pp.1-



6.doi:10.1109/INVENTIVE.2016.7823236.

URL: <http://ieeexplore.ieee.org/stamp/stamp.js?tp=&arnumber=7823236&isnumber=7823173>

- [15] Anupriya Koneru and Sreelatha M, “A Blind Key Signature Mechanism for Cloud Brokerage System,” journal of Advanced Research in Dynamic and Control Systems, Issue: 13-Special issue, 2018, pp. 770-776
- [16] Anupriya Koneru and Sreelatha M, “Broker decision verification system using MR cloud tree,” International Journal of Engineering & Technology, 7 (4) (2018) 3306-3311.
- [17] Koneru, Anupriya&Sreelatha, M.. (2017). Cloud service broker: Selection of providers using DTRFV evaluation. Journal of Theoretical and Applied Information Technology. 95. 3551-3559.
- [18] A. Koneru and M. Sreelatha, "A comprehensive study on cloud service brokering architecture," 2017 International Conference on Computing Methodologies and Communication (ICCMC), Erode, 2017, pp. 47-53. doi: 10.1109/ICCMC.2017.8282517 URL: <http://ieeexplore.ieee.org/stamp/stamp.js?tp=&arnumber=8282517&isnumber=8282515>
- [19] Purushottama Rao K., Koneru A., Naga Raju D. (2019) OEFC Algorithm—Sentiment Analysis on Goods and Service Tax System in India. In: Mallick P., Balas V., Bhoi A., Zobaa A. (eds) Cognitive Informatics and Soft Computing. Advances in Intelligent Systems and Computing, vol 768. Springer, Singapore.