WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

PREDICTION ACCURACY OF TERRORIST ATTACK USING ML

Prof. S. Premkumar [ASP], Monish V, Saravana Kumar P, Saravanan M

Knowledge Institute of Technology, Salem
Email ID: spece@kiot.ac.in, 2k18ece045@kiot.ac.in , 2k18ece065@kiot.ac.in , 2k18ece312@kiot.ac.in.

ABSTRACT

Terrorism has become one of the most tedious problem to deal with and a prominent threat to mankind. There are many identified problems, issues, and challenges in the terrorism studies that need to be addressed. The terrorist attack record has been stored in the form of database known as the "Global Terrorism Database (GTD)". The Global Terrorism Database has a total record of 140,000 terrorist incidents worldwide from 1970 to 2014 and has attributes namely timing, location, use of weapons and targets, number of casualties and responsible parties. Due to the rapid increase in terrorist activities throughout the world, there is serious intention required to deal with such activities. There must be a mechanism that can predict what kind of "attack types" can happen in future and important measures can be taken out accordingly. A hybrid classifier which combines different machine learning algorithms is proposed for the prediction of terrorists attack type. Hybrid classifier is designed using big data and it is proposed with some existing classifier such as K Nearest Neighbor, Support Vector Machine, Bagging and Decision Tree. The results reveal that hybrid classifier provides more accuracy than any single classifier in predicting terrorists attack type.

Keywords- *Global terrorism Database, Hybrid Classifier, Classifiers*

1. INTRODUCTION

Today's world is generating huge volumes of data at everaccelerating rates As a result big data analytic has become a powerful tool for businesses looking to leverage mountains of valuable data for profit and competitive advantage. Big Data is an emerging technology which allows to describes the process of extracting useful information from huge voluminous amount of structured data such as traditional databases, sensor values and GPS data, as well as unstructured data such as multimedia data, social Media's data and documents which are highly useful for mining the right information for our purpose which cannot possible in exiting traditional technology. And thorough the Big Data is possible Future predictions can be made easily and accurately. By analyzing huge volume of data anticipation can be made accurately based on the situation which highly beneficial for different sectors. Big data refers to datasets whose size is beyond the ability of typical database software tools to capture, store, manage, and analyze. This definition is intentionally subjective and incorporates a moving definition of how big a dataset needs to be in order to be considered big data i.e., we don't define big data in terms of being larger than a certain number of terabytes. With Big data huge amounts of data will be kept longer and have way more value than today's archived data at a lesser cost.

1.2 NEED OF PRESENT

1. Due to advancement of information transfer technologies the chances of possibility of getting attacked is high and amount of damage to property and people is huge in each attack cases
2. Prediction of possible attacks greatly reduces response time and increases chances to be prepared to face the adverse after-effects of an attack, thus decreasing the damage level

2. LITERATURE SURVEY

A. Introducing the Global Terrorism Database (Gary Lafree and Laura Dugan, 2007):

Compared to collecting data on other types of criminal violence, collecting data on terrorist violence is especially challenging. In the United States, the most widely used form of official crime data has long been the Federal Bureau of Investigation's Uniform Crime Report. First, terrorism data collected by government entities are suspicious either because they are influenced by political considerations, or because many fear that they might be so influenced. Availing political pressure is likely to be especially acute with regard to terrorism. Second, while huge amounts of detailed official data on common crimes are routinely produced by the various branches of the criminal justice system in most nations, this is rarely the case for terrorism. The GTD provide information about the frequency of attacks, and it allows us to examine the geographical distribution of terrorist tactics. By contrast, terrorist Attack

patterns for terrorist groups in Northern Ireland. Introducing the Global Terrorism Database 193 attacks in Asia and other regions relied less on bombs but were more likely to involve facility attacks. Finally, we see that in all regions of the globe, terrorists were less likely to rely on kidnappings and hijacking than on bombings, facility attacks, and assassinations. In response to the challenges raised by collecting valid data on terrorist events, researchers have been gradually developing more extensive open source terrorism data bases. At the moment, the Global Terrorism Database (GTD) is the largest and most extensive of these efforts.

B. Review of Group Prediction Model for Counter Terrorism Using CLOPE Algorithm (Pawan H. Pilley and S.S.Sikchi, 2014)

Prediction of terrorist group using historical data of attacks has been less explored due to the lack of detailed terrorist data which contain terrorist group's attacks and activities. The reasons may be its confidentiality & sensitivity. Current terrorism informatics, which aims to help security officials using data mining techniques, is mainly focused on using social network analysis (SNA) for structural and positional analysis of terrorist networks where required information is provided from non-crime data. Terrorist group prediction model (TGPM) which learns the pattern of terrorist attacks from the available historical data and make an association between terrorist group and previous attacks. Every terrorist group can be differentiated based on the style of attack, targets like police, private organizations; public property etc. so by analyzing these patterns TGPM will predict the group that may be involved in a given incident. GDM is a general detection model based on co-offending clustering. It works by linking co-offenders with inner join query using unique crime reference id numbers. CLOPE is specific to categorical attributes, which forms a large part of our database. At each level of the hierarchy; it groups instances within clusters, again with respect to a distance function. CLOPE attempts to increase the intra-cluster overlapping of categorical values by increasing the height-to-width ratio of the cluster histogram. It also uses a parameter called repulsion, to control the tightness of the cluster. Different number of clusters can be obtained by varying this parameter.

C. Using Global Terrorism Database (GTD) and Machine Learning Algorithms to Predict Terrorism and Threat (S. Kalaiarasi, Ankit Mehta, Devyash Bordia and Sanskar, 2018):

There has been a huge growth in Internet users in the past decade. Technological advancement has not only benefited the society but also has given rise to various problems in the society. One of such threats is the growth in cyber terrorism. Due to exponential increase in Internet users it is evident that people have started using this technological advancement for carrying out unlawful activities. It is very necessary to identify the fact that cyber terrorism has had very disastrous effect in our society. They finally went with kNN algorithm for Weapon Classifier and Random Forest algorithm for Perpetrator Classifier. K-NN algorithm was best suitable for Weapon Classifier and similarly Random Forest Algorithm stood out to be best for Perpetrator classifier on our multiple attempts using weapon classifier. The weapon classifier was built using kNN algorithm that classifies the attacks based on types of weapons. Thus, from the majority of the 12 attributes we predicted the weapons that could be used. The accuracy that we got using the k-NN algorithm was 88.74%. The perpetrator classifier classifies various groups or organization that carries out illegal and influences terrorism. We used Random Forest Algorithm for creating the Perpetrator Classifier. Hence, we got accuracy of 90.45%, %, and precision of 89.95% from our model using the Random Forest Algorithm for creating perpetrator classifier. Thus, the use of both methods can help to manage data sparsity problem and cold start problem in recommender system.

D. An Experimental Study of Classification Algorithms for Terrorism Prediction (Ghada M. Tolan and Omar S. Soliman, 2014)

Terrorist attacks are biggest, challenging, and leading issue in the whole world. There are various classification approaches proposed by the researchers in machine learning, statistics, and pattern recognition. This reviews the different data mining techniques that are being used for the classification and prediction and the prior work done on the respective topic. The techniques that are reviewed are Naïve Bayes, KNN, C4.5, ID3, and SVM. Naïve Bayes Classifier is the supervised machine learning technique used to take decision under the uncertain conditions as well as a statistical method for classification. K-Nearest Neighbor (KNN) Classifier is one of the top ten algorithms used for the classification and regression. ID3 is one of the popular DT algorithms that deal with nominal data sets, does not deal with missing values. Support Vector Machine (SVM) is a new and promising method for regression, classification, and general pattern recognition. SVM aims to find the best classification function to distinguish between members of the two classes in the training. In this approach we deal the missing data in our data set by using the mode and frequency distribution of the attributes to handle the missing data instances. The experiment conducted during the modeimputation approach, in case of test split of the input data with splits 66% for training data, and 34% for testing data showed that SVM is more accurate than other classifiers especially NB, and KNN, the overall performance of NB and KNN is almost the same. C4.5 has the lowest precision, recall, and F-measure in contrast with ID3 which has highest results in precision, recall, and F-measure although it is not accurate.

E. A Hybrid Classification Algorithms for Terrorism Prediction in Middle East and North Africa (Motaz M.H. Khorshid, Tarek H. M. AbouEl-Enie, Ghada M. A. Soliman, 2016)

Numerous machine learning methods and different knowledge representation models can be used to support decision making methods. The concept of combing classifiers is proposed as a new direction for the improvement of the performance of individual machine learning algorithms. The integration of the basic technologies into hybrid machine learning solutions facilitate more intelligent search and reasoning methods that match various domain knowledge with empirical data to solve advanced and complex problems. In this research study a real world data set of Middle East and North Africa is used for terrorism prediction based on hybrid machine learning algorithms with the help of WEKA as one of important machine learning software written in JAVA. Methods for learning comprehensible, human readable knowledge are especially appropriate in building knowledge based decision support systems/expert systems. Well known and famous methods are Decision Tree (DT), and rule Learning (RL), as well as Hoeffding Tree or Very Fast Decision Tree (VFDT), which is a new method introduced for incremental machine learning from data streams. The data set used in our research study is a real world data about terrorist events occurred in Middle East & North Africa in the period from 2009 till 2013, which consists of a total of 43335 terrorist events (instances), and 45 attributes, the attribute terrorist group is consisting of 120 diverse terrorist groups. Data reduction is performed on the terrorism data by selecting the most informative attributes that are highly correlated to our predicted attribute (Terrorist Group Name) without lose any

critical Information for classification and so 10 attributes are selected to be included in this experiment. For the missing data values, there are three approaches to handle missing data elements: removal, imputation, and special coding. Conducting the experiments to perform different classification algorithms on the research data set by using WEKA as one of important tools available for implementing data mining algorithms to train the base classifiers then the evaluation of the implemented classifiers is performed by using the testing data set. The overall results show that hybrid machine learning classifiers demonstrate good and proved obvious improvement in predictive accuracy over some standard comprehensible and ensemble methods.

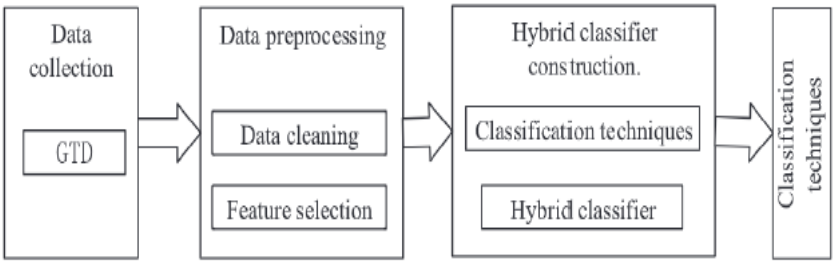
3. EXISTING SYSTEM

Terrorist attacks are the challenging issue across the world and need the attention of the practitioners to cope up deliberately. Predicting the responsible group of an event is a complicated task due to the lack of in depth terrorist historical data. It is evident that there has been enormous growth in terrorist attacks in recent years. The idea of online terrorism has also been growing its roots in the internet world. These types of activities have been growing along with the growth in internet technology. These types of events include social media threats such as hate speeches and comments provoking terror on social media platforms such as twitter, Facebook, etc. These activities must be prevented before it makes an impact. In the existing system they have made various classifiers that will group and predict various terrorism activities using k-NN algorithm and random forest algorithm. The main data source for this prediction of terrorists attack is taken from Global Terrorism Database (GTD). The Global Terrorism Database has a total record of 140,000 terrorist incidents worldwide from 1970 to 2014 and has attributes namely timing, location, use of weapons and targets, number of casualties and responsible parties. The terrorist attacks in GTD into transnational events, analyses and compares the impact of the both. There are many underlying correlations and laws in number associated with terrorist attacks, so it will be useful for further counterterrorism decision making and conduct regular identification and prediction. Terrorism has become a huge threat over the world. Various Machine learning system, artificial intelligence and Data-Analytics have provided with a system to help the investigator and anti-terrorist or counter-terrorist squad to rapidly decide the most probable perpetrator responsible of a particular terrorist attack. In the existing system they have explored various classifiers on the basis of their accuracy and speed. KNN algorithm for Weapon Classifier and Random Forest algorithm for Perpetrator Classifier. K-NN algorithm was best suitable for Weapon Classifier and similarly Random Forest Algorithm stood out to be best for Perpetrator classifier on our multiple attempts using various algorithms. The data-sets generated using Global Terrorism Database was divided into the ratio of 8:2 for training and testing the model respectively

4. PROPOSED SYSTEM

There are many identified problems, issues, and challenges in the terrorism studies that need to be addressed. Terrorism seems to be the most security challenge that is currently facing the world which requires urgent attention of researchers for possible solution. There are many possible approaches to the multifaceted problem of terrorist activities, such as the use of dialogue, arrest, detection, and gun which has not yield a positive result. There has been significant increase in terrorist attacks all over the world and this has become a serious issue of concern. The terrorist attack records have been stored in the form of database known as the “Global Terrorism Database (GTD). For Example, it has a record of 25,903 terrorist attacks in a span of twelve years between 2000 and 2012, which are approximately 2000 per year. In this proposed system, apart from using single classifier we go with hybrid classifier. There are two main types of Hybrid classifier: 1. the combination of centralized classification algorithm. 2. The combination of different classification algorithm by voting or superposition method. The reason to go with hybrid classifier is, the prediction accuracy of hybrid classifier is more than any single classifier in predicting “terrorist attack type”.

PROPOSED BLOCK DIAGRAM:



SOFTWARE REQUIREMENTS:

- Operating System : Windows 10
- Technology : Big Data Analytics
- Software used : Anaconda (Jupyter Notebook)
- Language : Python 3
- IDE : Python idle3

HARDWARE REQUIREMENTS:

- Processor: Intel Core i5-3210M
- CPU RAM: 4 GB
- Hard Disk: 500 GB

5. CONCLUSION

Terrorist attacks h a v e b e e n increasing in recent times, and prediction of these attacks is impossible. The GTD was used as big data, and hybrid classifiers were trained to predict new attacks. The data, after preprocessing for noise, was categorized based on the type of attacks. Then, the hybrid classifier with KNN, decision tree C4.5, bagging and SVM was designed. From the results, it is concluded that the accuracy of hybrid classifier is better than that of each single classifier.

REFERENCES

[1] Asma Gul, Aris Perperoglou, Zaradad Khan, Ensemble of a subset of KNN classifiers, advances in data analysis and classification December 2018, Volume 12, Issue 4, pp 827-840

[2] Sheikh H R. Use of Predictive modeling for prediction of future terrorist attacks in Pakistan [D]. 2016

[3] Soliman GMA, Abou-EI-Enien THM, Khorshid MMH. A comparison among support vector machine and other machine learning classification algorithms. Int J Comput Appl 2015:3-25

[4] GM Tolan, "An experimental study of classification algorithms for terrorism prediction

[5] G.Tolan,THM About-EI-Enien, MMH Khorshid, " hybrid classification algorithms for terrorism prediction in Middle East and North Africa, Iny J Emerg Trends Technol Comput Sci, 4(3)(2015),pp. 23-29

[6] S.Shafiq, W.H. Butt, U.Wamar, "Attack type prediction using hybrid classifier", International conference on advanced data mining and applications, 8933, Springer International Publishing (2014), pp. 488-498

[7] S, Nizamani, N.Menon, "Detecting terrorism incidence type from news summary", advanced information technology in education, 126, Springer Berlin Heidelberg (2012), pp. 25-102

[8] L.Rokach Ensemble-based classifiers Artif Intell Rev, 33(1- 2)(2010), pp. 1-39

[9] U Inyaem, P Meesad, C haruechaiyasak Named-entity techniques for terrorism event extraction and classification Eighth international symposium on natural language processing(2009),pp.175-179

[10] Lafree G. Dugen L. Introducing the global terrorism database. Terror political violence 2007; 19(2); 181-204