

Predictive Modeling of Terrorist Attacks Using Machine Learning

¹Chaman Verma, ²Sarika Malhotra, ³Sharmila and ⁴Vineeta Verma

¹Department of Media & Educational Informatics,

Faculty of Informatics,

EötvösLoránd University, Budapest, Hungary.

chaman.verma@gmail.com

²Imperial College of Engineering and Research,

JSPM, Wagholi, Pune, India.

sarika.malhotras19@gmail.com

³Manyawer Kashiram Rajkiya Polytechnic,

Tirwa, Kannauj, UP, India.

Kmsharmila@gmail.com

⁴Department of Basic Science,

Sardar Vallabhbhai Patel University of Agriculture and Technology,

Meerut, UP, India.

dr.vineeta.svp@gmail.com

Abstract

Machine learning algorithms play a vital role in prediction and classification of data in every domain. This paper presented three predictive models named attack type predictive (m1), attack region predictive (m2) and weapon type predictive (m3) which classify attack type, and attack region and weapon type based on millions of attacks using various supervised machine learning algorithms. The extracted data set is consisted of more than 0.17 million instances and 6 classes which are available online on the website of most popular dataset Global Terrorism Database (GTD) from National Consortium for the study of terrorism and Responses of Terrorism (START). The authors extracted only data set which contains information about terrorist attacks happened during the session 2013-2016 over the world. The classifiers support vector machine (SVM), Artificial Neural network (ANN), Naïve Bayes (NB), Random

Forest (RF), REP Tree and J48 are applied in Weka workbench. Further, the linear regression is also applied to find significant correlation between attacks and regression model is also evaluated by ANOVA test in R-Language. The findings of the study infer that RF performs better as compare to others to classify the attack type (84%) and attack region (100%) weapon type (91%). More than 70% True positive rate (TP rate) of Bombing/Explosion, Facility/Infrastructure attack, armed assault. The kappa statistic of m1, m2, and m3 are calculated 0.71, 1 and 0.82 prove the strong agreement among instances for accurate prediction. The linear regression model revealed the occurrence of Bombing/Explosion attack depends upon weapon type Explosives/Bombs/Dynamite. The positive correlation (0.65) is also found between weapon type and attack type.

Key Words:Accuracy, confusion matrix, kappa statistic, predictive, sensitivity.

1. Introduction and Related Work

Now a day, terrorism is the great problem for every nation in the world. Every citizen who is living wherever wants his or her security. This is the prime responsibility of every nation to protect the life of the citizen. In order to prevent this bad social evil, technology plays a vital role. Every country of the world is focusing on developing a preventive mechanism to avoid terrorist attacks. Hence, for prevention of terrorist attacks, predictive modeling is trending by various researchers. The terrorist attack prediction using supervised machine learning classifier is the conspicuous approach in data mining to generate predictive models. Hence, better data mining can be achieved either by supervised or unsupervised learning. In supervised learning, a data set is used to train by using some training model whereas in unsupervised learning technique no training set is used [1]. In the year 1970-1998, Hawkes Process is used to predict terrorist attacks in Northern Ireland which considered 5000 explosions [2]. Attacks are happening more and more nowadays, during the year 2013-2016, major 09 attack types in 205 countries over 12 regions with 22 targets using 12 weapon types were occurred [3]. To predict future attacks, machine learning is often used by many researchers in past. According to [4] random forest classifier (RF) has given 79% accuracy for attack types and for weapon type the accuracy of classification is 86% as compared to other classifiers. The social network analysis and pattern classification has been used to predict whether a person is terrorist or not and resulted in 86% accuracy [5]. SVM is more accurate than other classifiers especially NB, and KNN, the overall performance of NB and KNN is almost the same [6]. The crime prediction can also be made with group detection algorithms and CPM performed well on attributes of crime information to predict terrorist activities [7]. The terrorist group was predicted using combining various predictive models to achieve better accuracy [8]. More than 80% accuracy has been found by [9] to predict the terrorist group involved in a given attack in India from the year 1998 to 2008. The experimental study was also conducted on 43335 terrorist events by applying supervised machine learning classifiers which have proved SVM and RF gave better accuracy during classification [10].

2. Material and Methods

The experimental study is conducted on GTD dataset available on the website of National Consortium for the study of terrorism and Responses of Terrorism (START), University of Maryland USA, which contains millions of attack information of the world. The authors have used 170350 instances and 6 attributes. The authors extracted only data set which contains information about terrorist attacks happened during the session 2013-2016 over the world. The response attributes are attacked type, attack region and weapon type and rest of are the country, target and success. The list-wise deletion method is applied to handle the missing values in the dataset. The weapon attribute has 12 types of instances mentioned in table 2 and attack type attribute has 9 types of instances

(described in table 1). The attack region attribute has 12 types of instances (table 2). The authors have presented three predictive models (m1, m2, m3) to classify attack type, region type, and weapon type respectively. The performances of models are measures by true positive rate (TP rate), false positive rate (FP rate), Precision and recall. The agreement of attacks over dataset is tested by Cohen's kappa method. The six supervised machine learning classifiers are fitted on dataset using Weka 3.8.1 tool. The predictive models are presented after the successful comparison of accuracy with effective performance metrics. The Pearson product moment correlation is used to find a correlation between attacks and to predict attack type based on weapon type linear regression is also applied in R-language using a library(Hmisc). The significance of regression model is also evaluated by ANOVA test.

3. Experimental Environment

To present best predictive models as per objective, the section the section 3.1 explained predictive attack type model (m1) which analysis the prediction of various attack type. Subsequently, the section 3.2 explained classification of attack region to present predictive attack region model (m2) and later section 3.3 focused the predictive weapon type model (m3) which accurately predicts the weapon type. Section 3.4 proved the accurate prediction of attack type based on weapon type using linear regression, ANOVA in R- language.

Predictive Attack Type Model (m1)

The presented predictive model is fitted by Random forest supervised machine learning algorithm in Weka benchmark. The attack type is set as the response variable and remaining considered as independents or predictors. The accuracy of correctly classified instances is measured 84% and misclassification error is calculated 16% (Figure 1). The strong kappa statistic 0.7607 values proved the strong agreement among instances.

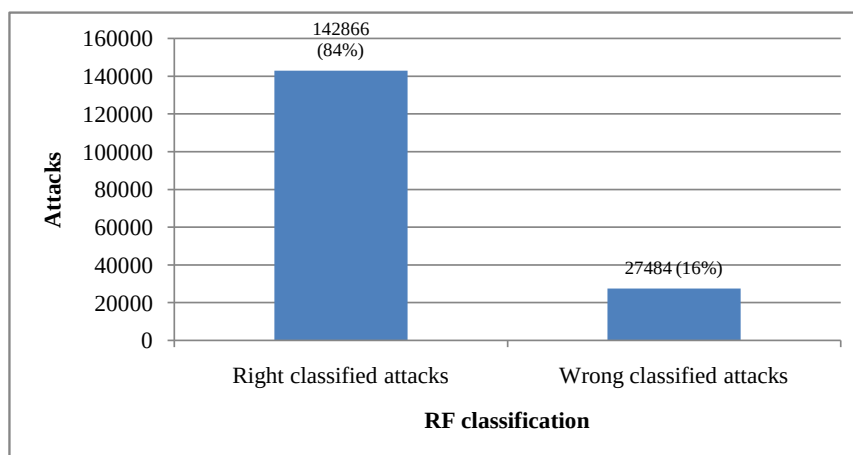


Figure 1: Attack Types Classification

Table 1 shows the parametric metrics of predictive attack type model to predict how accurately the model predicts the types of attack based on 5 independent variables discussed above in section 2. It can be seen that the correct positive prediction (TP rate/Recall/ Sensitivity) for class 3 (Bombing/Explosion) attack is 0.981 which accurately predicts higher attacks belongs to Bombing/Explosion class. Similarly, class 2, class 7, class 8, class 9 has higher TP rate which predicts more accurately attacks accordingly.

Table 1: Performance Metrics for Predictive Attack Type Model

TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
0.476	0.028	0.674	0.476	0.558	0.524	0.924	0.639	1 (Assassination)
0.854	0.098	0.729	0.854	0.787	0.718	0.953	0.846	2 (Armed Assault)
0.981	0.047	0.952	0.981	0.966	0.934	0.991	0.989	3 (Bombing/Explosion)
0.470	0.001	0.762	0.470	0.581	0.597	0.990	0.622	4 (Hijacking)
0.312	0.001	0.598	0.312	0.410	0.429	0.967	0.372	5 (Hostage Taking (Barricade Incident))
0.421	0.017	0.614	0.421	0.500	0.484	0.944	0.568	6 (Kidnapping)
0.818	0.009	0.845	0.818	0.831	0.821	0.990	0.897	7 (Facility/Infrastructure Attack)
0.675	0.001	0.799	0.675	0.732	0.733	0.999	0.846	8 (Unarmed Assault)
0.765	0.011	0.734	0.765	0.749	0.739	0.993	0.840	9 (Unknown)
0.839	0.051	0.831	0.839	0.830	0.793	0.972	0.876	Weighted Avg

The positive prediction (Precision) for Bombing/Explosion attack is found 0.952 which also states better performance of proposed model. For Facility/Infrastructure attack, the recall and precision are calculated as 0.818 and 0.845 respectively which infers large correct prediction of these attacks. The armed assault attack is also predicted correctly due to good TP rate (0.854) and precision (0.729). Further, model incorrectly classifies other attacks such as Assassination, Hijacking, Hostage Taking and Kidnapping.

Predictive Attack Region Model (m2)

In the classification of terrorist attack region, every classifier played their 100% role except only Naïve Bayes (NB) who missed 269 instances and ANN missed only 1 instance during the classification process. The attack region attributes have 12 type of instances shown in table 2.

Table 2: Attack Region

Region code	1	2	3	4	5	6	7	8	9	10	11	12
Name	North America	Central America & Caribbean	South America	East Asia	Southeast Asia	South Asia	Central Asia	Western Europe	Eastern Europe	The Middle East & North Africa	Sub-Saharan Africa	Australasia & Oceania

The predictive attack region model is found significant due to excellent kappa statistic which is 1 and means absolute error (MAE) is very low. The root means square error (RMSE) is also very low. The classification accuracy (CA) of all classifiers is 100% except ANN (99.9%) and NB (99.84%). The misclassification error (CE) is almost 0%.

Table 3: Performance Metrics for Predictive Region Type Model

	KS	MAE	RMSE	CA (%)	CE (%)
SVM	1.00	0.1389	0.2554	100%	0.00%
RF	1.00	0.0003	0.0036	100%	0.00%
REP Tree	1.00	0.00	0.00	100%	0.00%
J48	1.00	0.00	0.00	100%	0.00%
ANN	100.00%	0.00	0.0008	99.999%	0.00%
NB	0.9981	0.0021	0.0189	99.84%	0.16%

Predictive Weapon Type Model (m3)

RF model fitting on the dataset with weapon type as the response variable and remaining are predictors. The weapon class has 12 types of instances encoded in table 2. The average true positive rate (TP rate) is more than 90% which stated predictive weapon model is very meaningful for future. The presented weapon model is robust due to very good Cohen's kappa statistic 0.8497.

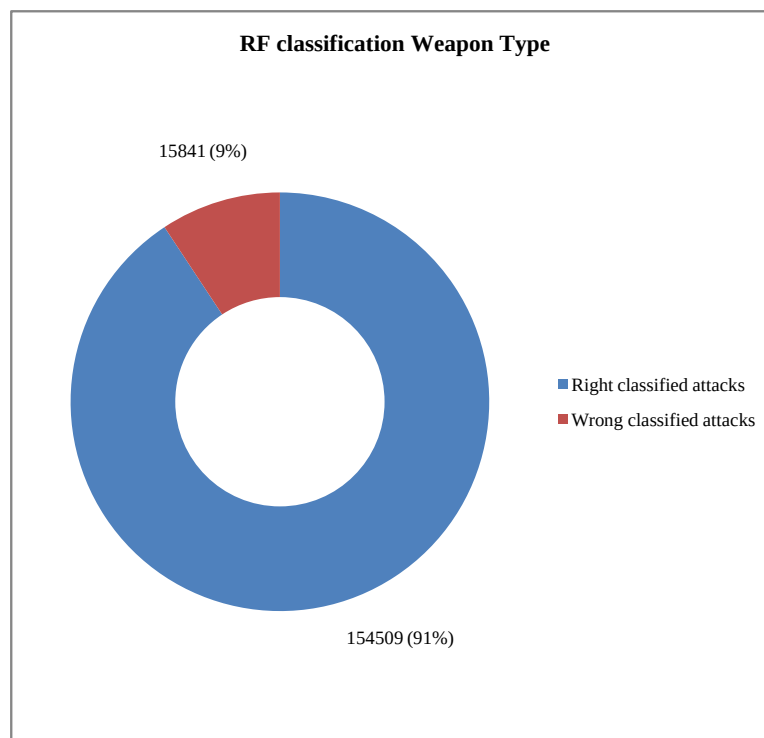


Figure 2: Weapon Type Classification

The Figure 2 shows that random forest (RF) given very high accuracy (91%) in predicting the weapon type in the data set. The number of accurate classified instances is 154509 out of 170350. The misclassification error is very low at 9%. Only 15841 instances are misclassified. Data from Table 4 shows predictive metrics for classify weapon types. The TP rate (1.00) of Radiological weapon predicts hundred percent of these weapons.

Table 4: Performance Metrics for Predictive Weapon Type Model

TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
0.800	0.000	0.757	0.800	0.778	0.778	1.000	0.824	1 (Biological)
0.611	0.000	0.895	0.611	0.726	0.739	0.994	0.740	2 (Chemical)
1.000	0.000	0.929	1.000	0.963	0.964	1.000	0.912	3 (Radiological)
0.926	0.085	0.840	0.926	0.881	0.822	0.974	0.939	5 (Firearms)
0.962	0.028	0.973	0.962	0.967	0.934	0.993	0.993	6 (Explosives/Bombs/Dynamite)
0.364	0.000	0.857	0.364	0.511	0.558	1.000	0.641	7 (Fake Weapons)
0.814	0.008	0.863	0.814	0.838	0.828	0.990	0.903	8 (Incendiary)
0.374	0.003	0.690	0.374	0.486	0.502	0.970	0.548	9 (Melee)
0.371	0.000	0.860	0.371	0.518	0.564	0.999	0.596	10 (Vehicle)
0.069	0.000	0.900	0.069	0.129	0.250	0.998	0.367	11 (Sabotage Equipment)
0.288	0.000	0.882	0.288	0.435	0.504	0.998	0.515	12 (Other)
0.710	0.012	0.845	0.710	0.772	0.757	0.981	0.879	13 (Unknown)
0.907	0.043	0.907	0.907	0.904	0.867	0.985	0.950	Weighted Avg.

Further, Firearms, Explosives/Bombs/Dynamite have more than 90% TP rate and Precision values which classify instances with higher accuracy. The sensitivity of Biological and Incendiary weapon is also more than 80% proves the model significance. The chemical weapon class has also good TP rate which also signifies presented model. Unfortunately, the model predicts less accurately class Fake Weapons, Melee, Vehicle and Other. Further, model incorrect classifies other attacks such as Assassination, Hijacking, Hostage Taking and Kidnapping.

Attack Correlation

In order to find a significant relation between six features, the authors used `rcorr()` function in the `Hmisc` package which yields significant correlation for Pearson and Spearman correlations methods. However, the input must be a matrix and pairwise deletion is used. The authors have also found the good correlation (0.65) between attack type and weapon type. The following lines of code are written in R Language for calculating the correlation between attack type and weapon type.

```
cor(mydata$`ATTACK TYPE`,mydata$`WEAPON TYPE`)
m1<- lm(ATTACK TYPE~WEAPON TYPE, data = dataset)
summary(m1)
plot(`ATTACK TYPE` ~ `WEAPON TYPE`, data=mydata)
abline(m1,col='red',lty=2,lwd=2)
library(Hmisc)
rcorr(mydata$`ATTACK TYPE`,mydata$`WEAPON TYPE`)
```

After explored significant correlation, the linear regression model is applied to predict attack type based upon weapon type from data using the following equation: $Y = |a| + |bX$ where Y is attack type, X is weapon type; b is the slope of the line and a is intercept of model. This equation is written as below:

```
model<-lm(mydata$`ATTACK TYPE`~ mydata$`WEAPON TYPE`, data = mydata)
```

Table 5: Regression Model of Attack Prediction

Coefficients:				
Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	-0.451543	0.010968	-41.17	<2e-16 ***
Weapon type	0.571161	0.001618	353.05	<2e-16 ***
Residual standard error: 1.437 on 170348 degrees of freedom				
Multiple R-squared: 0.4225, Adjusted R-squared: 0.4225				
F-statistic: 1.246e+05 on 1 and 170348 DF, p-value: < 2.2e-16				

Data from table 5 presents regression model summary of prediction of attack types using weapon type used by terrorist. The calculated intercept 0.57 shows the increment in the slope of the regression line for weapon type. The p-value for linear regression model is found $<2e-16$ *** which is significant. The residual standard error is found 1.43 which is proved fewer variations of attacks around the regression lines. The confidence interval for the model coefficient is 97.5%. Further, presented model is tested using ANOVA test which calculated the square root of mean 2 which is identified as the residual standard error of linear regression model of attack (table 5).

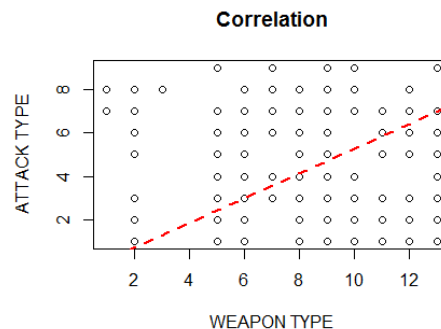


Figure 3: Regression Model of Attack Type Vs Weapon Type

Data from above figure 3 reflects positive correlation (0.65) between attack type and weapon type. In case the terrorist uses the Incendiary (8) and Fake Weapons (7) then the possibility of attack type is Hijacking (4). If the weapon type is Explosives/Bombs/Dynamite (6), the probability of attack type is near to Bombing/Explosion (3) which seems quite a logical prediction. The following command is used to predict attack type based on weapon type from the dataset.

```
head(predict(m1,data.frame("WEAPON TYPE"=6)))
```

In order to calculate the fit predicted values for the model, following line of code is used:

```
(predict(m1,interval = "prediction"))
```

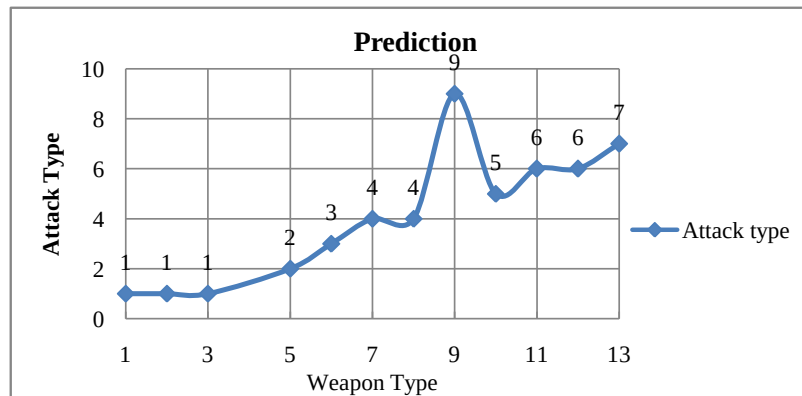


Figure 4: Predictive Values for Attack Types

Figure 4 presents the accurate prediction of attack type based on the weapon type provided by the regression model. In case, terrorist use any unknown weapon (13) then facility/Infrastructure (7) attack might happen. In case of usage of another category of weapon (12) and sabotage equipment (11), then Explosives/Bombs/Dynamite (6) attack can happen. For weapon firearms (5), armed assault (2) attack may occur. For weapon categories biological (1), chemical (2) and radiological (3), the model predicting attack type of assassination (1).

4. Discussion

The authors have analyzed the performance of attack predictive models in the previous section. It can be seen that predictive attack type model (m1) used with RF achieved 84% accuracy to predict the response variable named attack type. The accuracy is same gained by both classifiers RT and J48. The SVM outperformed the ANN and NB in terms of accuracy. The lowest accuracy is achieved by ANN classifiers.

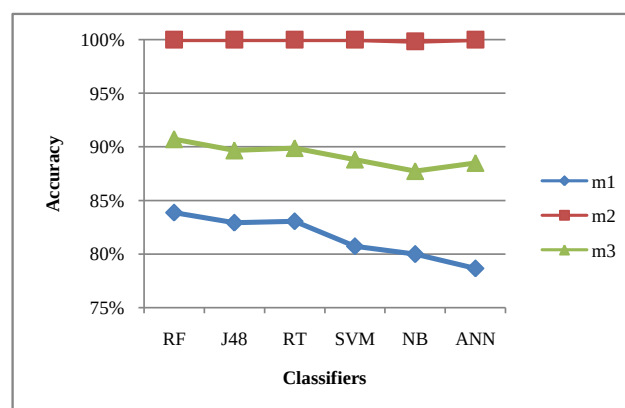


Figure 5: Accuracy Vs Classifiers

Data from Figure 5 reflects each classifier have achieved more than 75% accuracy which reveals the significance of every model. In case of attack region prediction, every classifier provided 100% accuracy which stated predictive region model (m2) is the best model to use in future. Therefore, attack region can be easily predicts based on the selected predictors. The model (m3) achieved 91% accuracy using RF and 90% using RT and J48 classifiers. The SVM classifier also proved better than NB and ANN for weapon type prediction. Hence, the predictive weapon type model (m3) is also proved better for prediction of weapon used by terrorist. The model (m1) predicted attack type more accurately with the support of RF with 84% accuracy. The classifiers J48 and RT have achieved the same accuracy (83%) which is higher than SVM (81%), NB (80%) and ANN (79%).

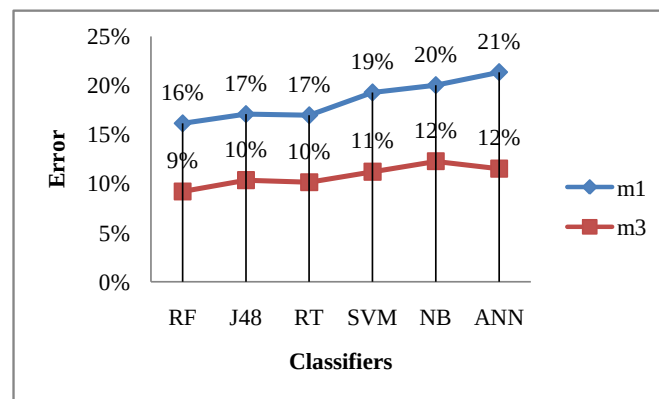


Figure 6: Error Vs Classifiers

It can be seen from Figure 6 that the error rate of each classifier is found lesser than 21%. As we have mentioned in section 3.2 that each classifier in region predictive model (m3) achieved almost 100% except NB classifier. The model (m3) has very less error rate 9% at RF classifier which infers the better prediction of weapon type. Further, J48 and RT classifiers have the same error rate 10% and higher misclassification error is achieved by ANN and NB for weapon prediction.

5. Conclusion

This experimental study is conducted in order to predict terrorist attacks from historical data available on START. The authors have presented three predictive attacks models m1, m2 and m3 for attack type, region type, and weapon type respectively. These predictive models have been fitted with most popular supervised machine learning algorithms such as RF, J48, RT, SVM, NB and ANN for the classification of attacks. Further, in comparison of classification accuracy, RF outperformed than other classifiers for three models. For the model (m1 and m3), the J48 classifier achieved higher accuracy (90%) than SVM, NB, and ANN. In order to predict region (m2), all classifiers have

achieved 100% accuracy. It is also proved that SVM classifier outperformed than ANN and NB in classification accuracy for both of attributes weapon type (89%) and attack type (81%) with leading nature of NB (79%) classifier over ANN (79%) classifier for attack type. These outcomes of the study are also supporting earlier study [6]. Hence, it is proved that RF achieved higher accuracy 84% attack type, 91% for weapon type and 100% for attack region which significant improvement of a study conducted by [4]. After the comparing the accuracy of RF classifier with the accuracy of others, the authors described important performance metrics for attack type and weapon type. The Cohen kappa statistic of all models found very good ($m1=0.71$, $m2=1$, $m3=0.82$) proves the strong agreement among instances for accurately attack prediction. On the basis of high precision value (table 1, table 4), the maximum accurate classification of Bombing/Explosion attack and Explosives/Bombs/Dynamite weapon. Further, the linear regression model proved significantly the occurrence of Bombing/Explosion attack if the type of weapon is Explosives/Bombs/Dynamite leads to meaningful prediction. The positive correlation has been found between attack type and weapon type. On the basis of weapon categories biological, chemical and radiological, regression model predicting attack type of assassination. The facility/Infrastructure attack may be happening if they use unknown weapon type.

Declaration

Availability of Data and Material

The dataset is available online on the website of National Consortium for the study of terrorism and Responses to Terrorism (START).

Competing Interests

The authors declare that they have no competing interests.

Funding

This research study is not funded by any institution or industry.

Acknowledgment

The authors would like to thank National Consortium for the study of terrorism and Responses to Terrorism (START) to provide this data online.

References

- [1] SA S., Intelligent heart disease prediction system using data mining techniques, Int J Healthcare Biomed Res 1 (2013), 94-101.
- [2] Swanson Wonkblog A., The eerie math that could predict terrorist attacks (2016).

- [3] Global Terrorism Database (GTD), <http://www.start.umd.edu/gtd>, 2017.
- [4] Saha S. et.al., Future Terrorist Attack Prediction using Machine Learning Techniques (2017).
https://www.researchgate.net/publication/317032840_Future_Terrorist_Attack_Prediction_using_MachineLearning_Techniques, Accessed on 1st April 2018.
- [5] Coffman T.R., Marcus S.E., Pattern classification in social network analysis: A case study, IEEE proceedings. Aerospace conference 5 (2004), 3162-3175.
- [6] Tolan G.M., Soliman O.S., An Experimental Study of Classification Algorithms for Terrorism Prediction, International Journal of Knowledge Engineering 1(2) (2015), 107-112.
- [7] Ozgul F., Erdem Z., Bowerman C., Prediction of unsolved terrorist attacks using group detection algorithm, Pacific-Asia Workshop on Intelligence and Security Informatics (2009), 25-30.
- [8] Faryal G., Wasi B.H., Usman Q., Terrorist group prediction using data classification, Proceedings of the International Conferences of Artificial Intelligence and Pattern Recognition, Malaysia (2014).
- [9] Sachan A., Roy D., TGPM: Terrorist Group Prediction Model for Counter-Terrorism, International Journal of Computer Applications 44(10) (2012), 49-52.
- [10] Khorshid M.M., Abou-El-Enien T.H., Soliman, G.M., Hybrid Classification Algorithms For Terrorism Prediction In Middle East And North Africa, International Journal of Emerging Trends & Technology in Computer Science 4(3) (2015), 23-29.

