Hindawi
Scientific Programming
Volume 2021, Article ID 7890923, 15 pages
https://doi.org/10.1155/2021/7890923

Research Article

Quantitative Analysis and Prediction of Global Terrorist Attacks Based on Machine Learning

Xiaohui Pan [1,2]

[1]*School of Artificial Intelligence and Law, Shanghai University of Political Science and Law, Shanghai 201701, China*
[2]*School of Management, Shanghai University, Shanghai 200444, China*

Correspondence should be addressed to Xiaohui Pan; panxiaohui@shupl.edu.cn

Received 15 May 2021; Revised 1 September 2021; Accepted 15 September 2021; Published 27 September 2021

Academic Editor: Wei-Chuen Yau

Copyright © 2021 Xiaohui Pan. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Terrorist attacks pose a great threat to global security, and their analysis and prediction are imperative. Considering the high frequency of terrorist attacks and the inherent difficulty in finding related terrorist organizations, we propose a classification framework based on ensemble learning for classifying and predicting terrorist organizations. The framework includes data preprocessing, data splitting, five classifier prediction models, and model evaluation. Based on a quantitative statistical analysis of terrorist organization activities in GTD from 1970 to 2017 and feature selection using the SelectKBest method in scikit learn, we constructed five classification and prediction models of terrorist organizations, namely, decision tree, bagging, random forest, extra tree, and XGBoost, and utilized a 10-fold cross-validation method to verify the performance and stability of the proposed model. Experimental results showed that the five models achieved excellent performance. The XGBoost and random forest models achieved the best accuracies (97.16% and 96.82%, respectively) of predicting 32 terrorist organizations with the highest attack frequencies. The proposed classifier framework is useful for the accurate and efficient prediction of terrorist organizations responsible for attacks and can be extended to predict all terrorist organizations.

1. Introduction

Terrorism is a complex political and social phenomenon. Terrorist attacks have a significant threat to the safety and security of the international community and have become one of the greatest obstacles to the sustainable development of global social security. Antiterrorism is an important part of global security governance, which is a sustainability issue that guarantees global security development. At present, terrorist attacks occur frequently, which leads to significant threats and poses a challenge to global social security governance [1]. According to statistics from the Global Terrorism Database (GTD) [2], more than 200,000 terrorist attacks have been recorded from 1970 to the present day. Terrorist attacks typically involve high lethality and destructive power and directly cause massive casualties and property losses. In addition, they bring tremendous psychological pressure on people. In summary, terrorist attacks result in social unrest to a certain extent, obstructing the regular order of work and life and thus greatly hindering economic development.

The analysis and prediction of terrorist attacks support targeted attacks on terrorist groups and provide valuable information for antiterrorism and terrorism prevention operations, enabling authorities to find new or hidden terrorists as soon as possible to reduce human and property losses, prevent problems and improve the security and stability of social life. The patterns of attacks planned and carried out by terrorists may seem random on the surface, but in fact, they are typically organized and premeditated actions chosen carefully and deliberately. Moreover, attacks by the same organizations and individuals tend to be substantially related in terms of certain distinguishable characteristics. Therefore, there must be some patterns or informal rules guiding the activities of terrorist organizations. After analyzing these characteristic patterns of activity by terrorist organizations, authorities can make more detailed predictions and analyses of terrorist organizations to

attack them more accurately and increase the time available for the prevention and prediction of terrorist attacks. The GTD provides researchers with comprehensive, reliable, and open-source data, in which there are many potential correlations and patterns to be found. Mining and identifying these patterns using digitally driven methods has become a research topic in the field of informatics.

During the past decades, scholars have established valuable models and algorithms for early warning and prediction of terrorist attacks. Ding et al. [3] demonstrated a novel method using relatively popular and robust machine learning methods to simulate the risk of terrorist attacks at a global scale based on multiple resources, long time series, and globally distributed datasets. The model performed relatively well in predicting the places where terror events might occur in 2015. Chuang et al. [4] studied the spatiotemporal patterns of terrorist attacks by Al Qaeda (AQ), the Islamic State of Iraq and Syria (ISIS), and various local militias or insurgents by applying data-driven, unsupervised k-means clustering to the GTD. Petroff et al. [5] proposed a hidden Markov model to generate early warnings of specific terrorist attacks. Gohar et al. [6] proposed a new collection framework for classifying and predicting terrorist organizations. The framework consists of four basic classifiers, including naive Bayes (NB), k-nearest neighbor (KNN), Iterative Dichotomiser 3 (ID3), and a decision stump (DS). Tolan et al. [7] employed classification techniques to compare five basic classifiers, including naive Bayes, NB, KNN, and support vector machine (SVM), and utilized the GTD to study terrorism and terrorist reactions. Meng et al. [8] proposed an optimized hybrid classifier including data collection, preprocessing, hybrid classification, mining, and classifier testing as a framework for terrorist attack prediction. Bu et al. [9] combined an SVM with an intelligent tuned harmony search (ITHS) algorithm to build an ITHS-SVM model for terrorist attack classification, which provides learning and curve-fitting functions while optimizing SVM parameters. Li et al. [10] proposed a comprehensive framework that combined social network analysis, wavelet transform, and pattern recognition approaches to investigate the dynamics and eventually predict the attack behavior of terrorist groups. Hu et al. [11] developed a risk assessment system for terrorist attacks through a quantitative analysis of the GTD and clustered and ranked terrorist attacks according to the results of terrorist attack rating models. Campedelli et al. [12] proposed the use of temporal metagraphs and deep learning to forecast future terrorist targets using real data of attacks in Afghanistan and Iraq from 2001 to 2018. The experimental results showed that bidirectional LSTM networks achieve superior forecasting performance compared to other algorithms. Although these existing models and algorithms can be used to classify and predict terrorist activities, their accuracy remains less than ideal, and the coverage of the number of terrorist organizations is insufficiently comprehensive, with some covering only a few years or a few regions of a country.

In addition, methods designed to perform analysis and prediction of terrorist attacks have been gradually developed from the perspective of network science. Campedelli [13] proposed a new methodological framework integrating network science, mathematical modeling, and deep learning to compare and analyze the world's most active jihadist terrorist organizations (i.e., the Islamic State, the Taliban, AQ, Boko Haram, and Al-Shabaab), investigate their behavioral patterns, and forecast their future actions. Campedelli et al. [14] first built a multiparty network including information about terrorist organizations and tactics, weapons, targets, and active areas by using GTD data from 1997 to 2016. Then, a new clustering algorithm was proposed, which used von Neumann entropy for modal weighting. Compared with the other two clustering methods, the experimental results showed that the entropy-based method tended to reliably reflect the data structure naturally generated by the baseline Gower method. Desmarais and Cranmer [15] constructed a network of transnational terrorist attacks in which the source (sender) and target (receiver) countries share a directed edge. A deterministic, similarity-based link prediction framework was integrated into a probabilistic modeling approach to develop an edge-forecasting method. Experiments showed that probabilistic link prediction could not only accurately predict terrorist actions but showed promise to predict the onset of terrorist hostilities between a source and a target.

In recent years, artificial intelligence has rapidly emerged and gradually matured, empowering scientific and technological products, and promoting its development in human society. Artificial intelligence (machine learning) focuses on data training and fitting. With the support of a large amount of raw data, models can achieve a high prediction accuracy. Specific patterns may be noted in the organization, planning, and development of terrorist attacks, and the accumulated terrorist attacks over the years provide a large amount of characteristic data. Therefore, artificial intelligence is expected to become an excellent tool for analyzing and predicting the rules of terrorist attacks. Guo et al. [16] summarized three ways to improve conflict forecasting and called on the UN to invest in data-driven predictive methods for promoting peace. Three new methods were developed, including new machine learning techniques, more information on the wider causes of conflicts and their resolution, and theoretical models that better reflect the complexity of social interactions and human decision-making.

However, well-known machine learning problems, such as data imbalance, the curse of dimensionality, and false correlation, may cause machine learning algorithms to be inaccurate. When using datasets with a sufficient number of labeled cases, machine learning can help police departments detect local crimes and predict when and where crimes will occur. However, when predicting the identity of offenders or criminals, especially in terrorist attacks, the number of false positives and false negatives can be relatively high when implementing these algorithms because the feature attributes are numerous and redundant.

The present work is focused on the application of machine learning to study the characteristics of terrorist attacks, quantitative analysis, and prediction. There are many types of artificial intelligence models and methods, and the

ensemble learning method in machine learning has higher accuracy and better generalization ability. In this study, we perform a quantitative analysis of the activities of terrorist organizations in GTD from 1970 to 2017 and propose a classification framework based on ensemble learning to accurately classify and predict terrorist organizations. First, we perform a quantitative analysis of global terrorist attacks and study the varying attributes and characteristics of different times, places, and terrorist organizations in the GTD. Second, we perform essential data preprocessing, including data cleaning (such as missing value processing and data conversion), feature selection, and data splitting. Third, we construct a classification framework based on ensemble learning for classifying and predicting terrorist organizations. The framework incorporates five current mainstream models: decision tree, bagging, random forest, extra tree, and XGBoost. Finally, we conduct comprehensive experiments to evaluate the performance of these algorithms through a set of metrics and provide a visual exploratory discussion. Experimental results show that the XGBoost and the random forest models were the most effective and achieved the highest accuracy.

2. Analysis of GTD Dataset

2.1. Quantitative Analysis of GTD Dataset. The dataset was derived from data on terrorist attacks from 1970 to 2017 from the GTD [2] (https://www.start.umd.edu/gtd/), which is managed by the National Consortium for the Study of Terrorism and Responses to Terrorism (START). The GTD dataset is considered to be the most comprehensive database for recording global terrorist activity. The information of terrorist organizations in GTD is represented by the "gname," "gname2," and "gname3" attribute fields, which, respectively, represent up to three organizations participating in an event. Most events have only a gname field value, and some events may only be represented by an unknown. Therefore, this article is focused on analyzing and predicting the "gname" attribute field (that is, the major organization). For a very small number of events with more than one terrorist organization, we focus on the major terrorist organization. According to the analysis of the dataset, there were 3,537 nonrepeated statistics on the attribute fields of the terrorist organizations recorded. Except for the records of unknown terrorist groups as "gname = Unknown," there were 3,536 terrorist groups in the dataset.

Our preliminary analysis shows that there were 181,691 identified terrorist incidents in the GTD dataset from 1970 to 2017, excluding incidents with unknown terrorist organizations. Among the cases where terrorist groups have been identified, some terrorist organizations were very active and launched numerous attacks; 19 large terrorist organizations that launched more than 1,000 terrorist attacks, 32 terrorist organizations exceeded 500, and 122 terrorist organizations exceeded 100. These 122 terrorist organizations launched 78,107 terrorist attacks, accounting for more than 79% of all known terrorist group incidents. The number of terrorist attacks, the corresponding number of terrorist organizations, and the sum of terrorist attacks by these terrorist organizations are listed in Table 1.

The detailed statistics of terrorist attacks and the corresponding number of terrorist organizations are shown in Figure 1. Among them, the areas within the rectangles of each specified range were mutually exclusive.

2.2. Visual Analysis of Terrorist Organizations. The activities of the 3,536 terrorist organizations were further analyzed. To facilitate visual observation, we conducted a detailed analysis of 32 terrorist organizations that carried out more than 500 terrorist attacks from 1970 to 2017. The names, number of attacks, and ranking of these 32 terrorist organizations are shown in Figure 2. The organization with the maximum number of terrorist attacks was the Taliban, which conducted a total of 7,478 terrorist attacks, followed by the Islamic State of Iraq and the Levant (ISIL) and Shining Path (SL) with 5,613 and 4,555 terrorist attacks, respectively. In addition, the number of terrorist attacks slowly decreased from around 3,000 (FMLN) to around 500 (Fulani extremists).

The annual activity distribution of the top 10 terrorist organizations was calculated, as shown in Figure 3. Notably, the overall number of terrorist attacks has shown a significant upward trend in recent years. In particular, the three terrorist organizations ISIL (purple line), Taliban (light blue line), and Al-Shabaab (blue line) planned and carried out terrorist attacks extremely frequently. It is worth noting that ISIL emerged only in 2012 and has since launched more than 1,000 terrorist attacks every year. The intrinsic reason is worthy of in-depth analysis by researchers.

3. Research Methods

In this paper, we propose a classification framework based on ensemble learning to classify and predict terrorist organizations. The framework involved four steps, including data preprocessing, data splitting, construction and training of several ensemble learning classifier models, and classifier model testing, as shown in Figure 4.

3.1. Machine Learning Prediction Model. Machine learning methods are usually divided into supervised and unsupervised approaches. Within the former category, many applications aim at predicting a target variable. The specific method involves establishing a corresponding relationship between the attribute variables and the target variables in the sample dataset, and this mapping relationship is formed by constructing a model from the training dataset. The prediction and evaluation are performed on the testing dataset, and the value of the predicted target variables is then compared with the value of real target variables to derive the prediction accuracy.

For a given terrorist attack, the classification models identify terrorist organizations or individuals in a terrorist attack based on known attribute fields. In the process of supervised machine learning, the existing terrorist event feature data are sent to the classification algorithm model for

Table 1: Statistics on the number of terrorist attacks and terrorist organizations.

Number of terrorist attacks	Number of terrorist attack organizations	Sum of terrorist attacks
≥1000	19	50200
≥500	32	58520
≥100	122	78107
≥50	210	84339
≥5	936	94871

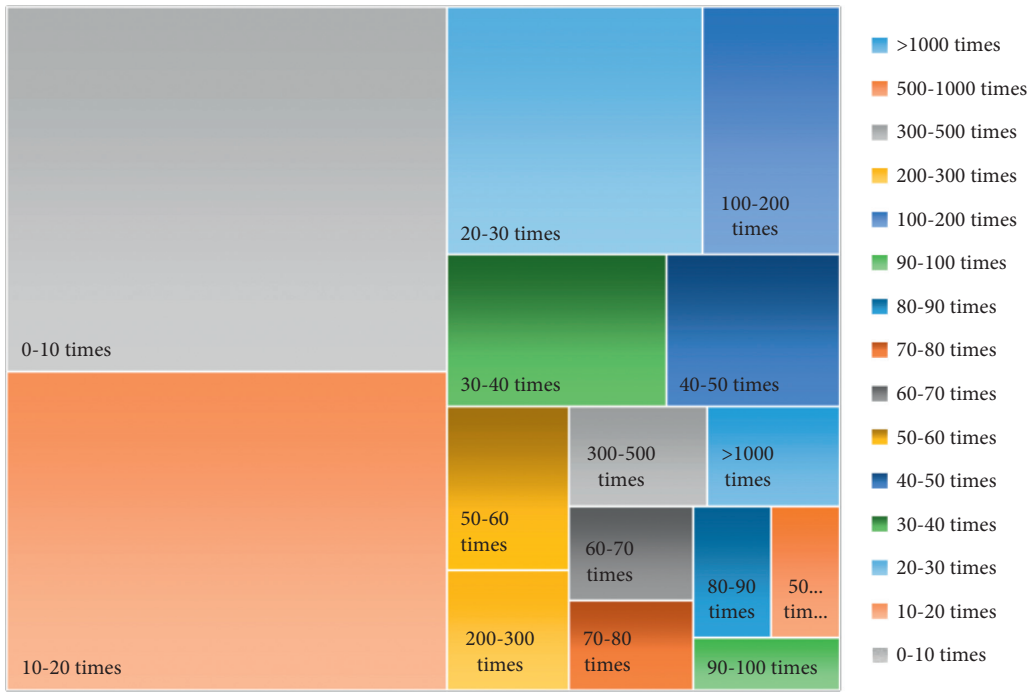

Figure 1: The statistics on the number of terrorist organizations by the number of attacks.

training and learning. Then, the trained model is used to classify the test or new data to predict candidate terrorist organizations or individuals. Therefore, the prediction of terrorist organizations is a multiclassification problem. The primary purpose of this research is to construct classification models for multiclassification tasks. In this study, we used five supervised machine learning classifiers [17] to predict terrorist organizations responsible for various attacks, including decision tree, bootstrap aggregating, random forest, extra trees, and super gradient boost.

Decision tree (DT) algorithms [18, 19] can be used as a supervised learning method. By creating a tree model to learn simple decision rules from data features to predict the value of a target variable, the DT model begins the decision from the root node, and the leaf nodes represent a successful guess or correct prediction. There are three major algorithms for creating DTs: ID3, C4.5, and Classification and Regression Tree (CART). ID3 starts from the root node of the tree and uses information gain to select features to build child nodes. C4.5 uses the information gain ratio to select features, which is regarded as an improvement of ID3. However, these two algorithms cause the problem of overfitting, which requires pruning. The pruning of the DT removes unnecessary classification features by optimizing the loss function and reducing the overall complexity of the model. CART [20, 21] adopts the Gini index minimization principle to create a tree. It cuts out some subtrees from the bottom of a fully grown DT, making the model simpler. We used CART to create decision trees in this study.

The following four models are considered ensemble learning [22], which is a branch of machine learning. The basic unit of these four models is a decision tree.

Bootstrap aggregating (Bagging) [23, 24] is a classification algorithm that uses a combination strategy. It first obtains m sample sets by extracting the original dataset m times with replacement and then uses each sample set to train m base classifiers separately. Finally, an integrated classifier was constructed by applying a combination strategy to the base classifiers.

Random forest (RF) [25] is an algorithm that integrates multiple DTs through ensemble learning. RF usually uses the mean or mode of the prediction results of each DT in the decision tree set as the final prediction value. The RF in the scikit-learn Python package uses the mean as a predictor. Compared with a single DT, RF is less likely to be affected by overfitting because each DT of the random forest cannot see

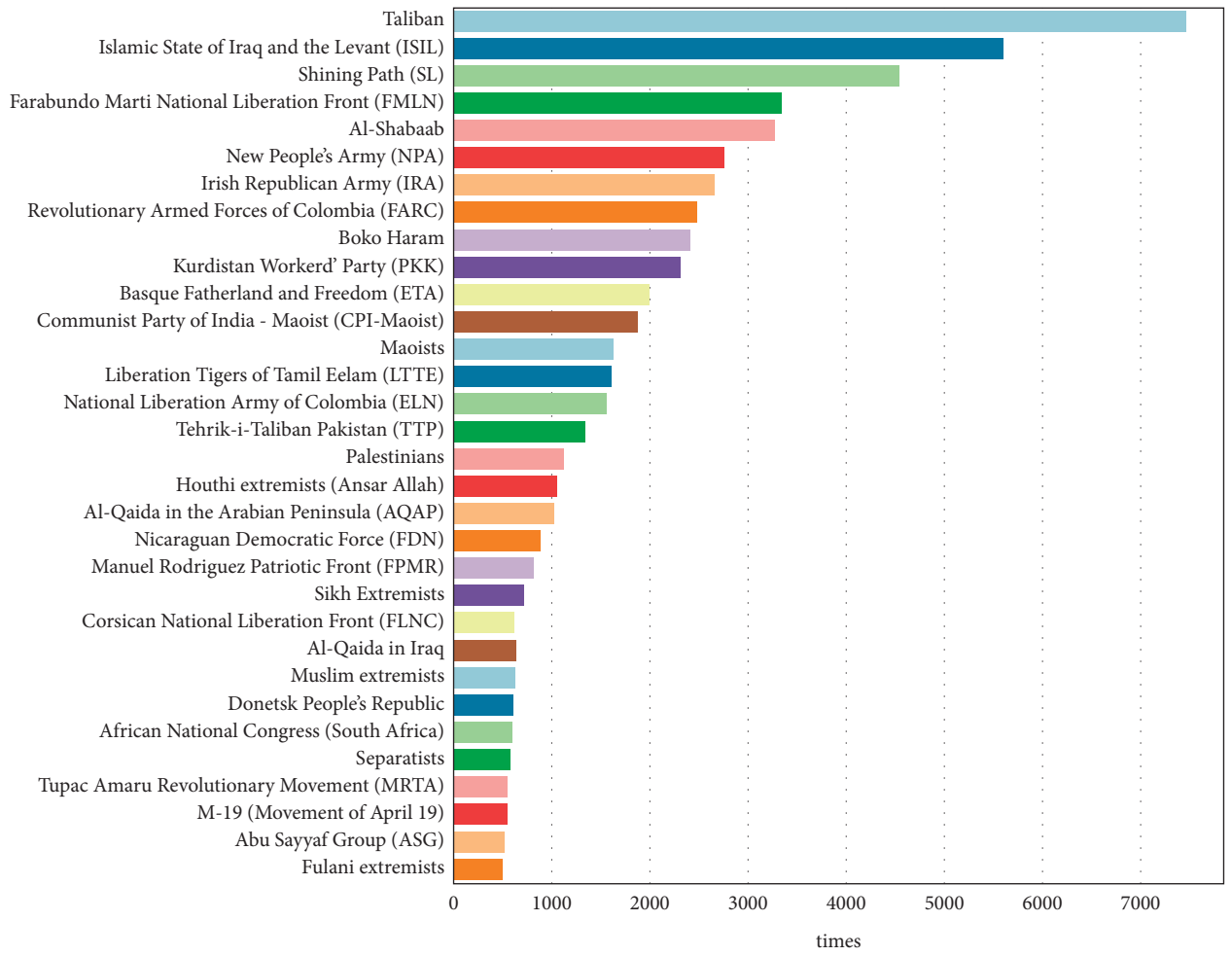

Figure 2: Rankings of terrorist organizations with more than 500 terrorist attacks.

the full view of the training set. Each DT only trained a part of the attribute data and did not remember all the noise of the training set.

Extra trees (ET) [26, 27] are also composed of many DTs, such as RF. These decision trees use random features and random thresholds for the node division. ET provides additional randomness, which suppresses overfitting but also increases the bias to some extent. The difference between ET and RF is that RF uses bagging for random sampling, whereas ET uses all samples. RF finds the optimal attributes based on information entropy and the Gini index in a random subset, while ET finds an eigenvalue entirely at random to divide.

The super gradient boost (XGBoost) [28, 29] is also a classification algorithm that integrates multiple decision trees. It pays more attention to the samples that were learned incorrectly in the previous round during training and makes some improvements on Gradient Boosting by introducing second-order derivatives and approximating the loss function with first- and second-order derivatives so that there is more information in the optimization process. In addition, XGBoost adds a regular term to the loss function to weigh the complexity of the model, making it simpler and preventing overfitting. Compared with the RF, there is no dependency relationship between the decision trees in the RF, and they can be parallel. However, XGBoost trees are dependent and must be serialized. This model maximizes the integration speed and efficiency of trees and is a very effective integration algorithm.

3.2. Evaluation Metrics. After the machine learning classification model for this problem was designed and constructed, it was necessary to evaluate the performance of a classifier to determine the accuracy of a classifier in predicting the class labels of terrorist organizations.

In machine learning, a multiclass classification problem can usually be converted into multiple binary classification problems. Each binary classification problem classifies a group of target objects into one class (i.e., category) and the remaining target objects into another class. The confusion matrix is an analysis table that summarizes the prediction results and the real results in binary classification and multiclass classification [30], as shown in Table 2.

Based on confusion matrices, four commonly used metrics are generally applied to evaluate the performance of machine learning, including accuracy, precision, recall, and $F1$ score.

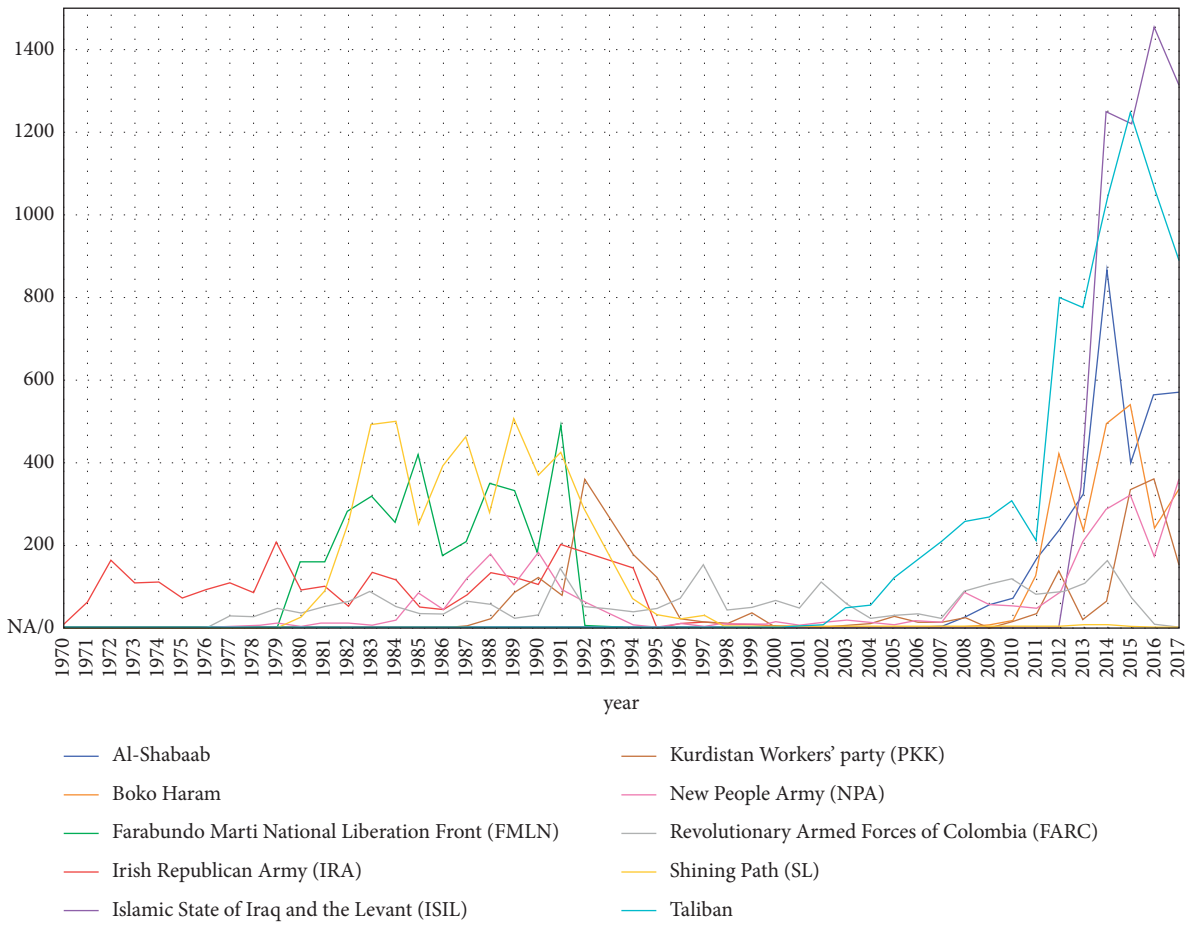

Figure 3: The annual activity distribution of the top 10 terrorist organizations.

Among them, accuracy can intuitively reflect the prediction result. Precision and recall are good complementary indicators when accuracy is not sufficient to reflect the detail of the assessment results. The $F1$ score is the harmonic mean value of the precision and recall. According to Table 2, the four metrics are defined as follows:

$$\begin{aligned} \text{Accuracy} &= \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}, \\ \text{Precision} &= \frac{\text{TP}}{\text{TP} + \text{FP}}, \\ \text{Recall} &= \frac{\text{TP}}{\text{TP} + \text{FN}}, \\ F1 &= \frac{2 * \text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}}. \end{aligned} \tag{1}$$

Accuracy is defined as the ratio of correctly predicted samples to the total number of samples. It is the percentage of terrorist organizations correctly classified in an attack. Precision is the ratio of true positive samples among all samples predicted as positive samples. Recall is the ratio of the number of positive samples predicted to the total number of all positive samples. For specific terrorist organization i, precision (abbreviated as P) refers to the ratio of the number of samples correctly predicted to be terrorist organization i to the number of all samples predicted to be terrorist organization i. Recall (R) refers to the ratio of the number of samples correctly predicted to be terrorist organization i to the number of true samples of a terrorist organization i.

P and R indicators are important to provide an alternative perspective on the false positive rate versus the false negative rate of that terrorist organization. A false negative means that the organization launched an attack that was not correctly identified and instead confused with another terrorist organization. A false positive means that an attack that the group did not initiate was incorrectly identified as its act. Comparing the importance of false positives and false negatives for a terrorist organization is different for different terrorist organizations. Because of the impact this can have on the subsequent handling strategy of a terrorist attack situation, for larger terrorist organizations with more aggressive activities and more severe tactics, false negatives are more important and have more serious consequences, as they may result in negligence and trivialization of the handling strategy. For less aggressive, relatively civilized terrorist organizations, false positives are more important and may likewise result in negligent and dismissive treatment strategies.

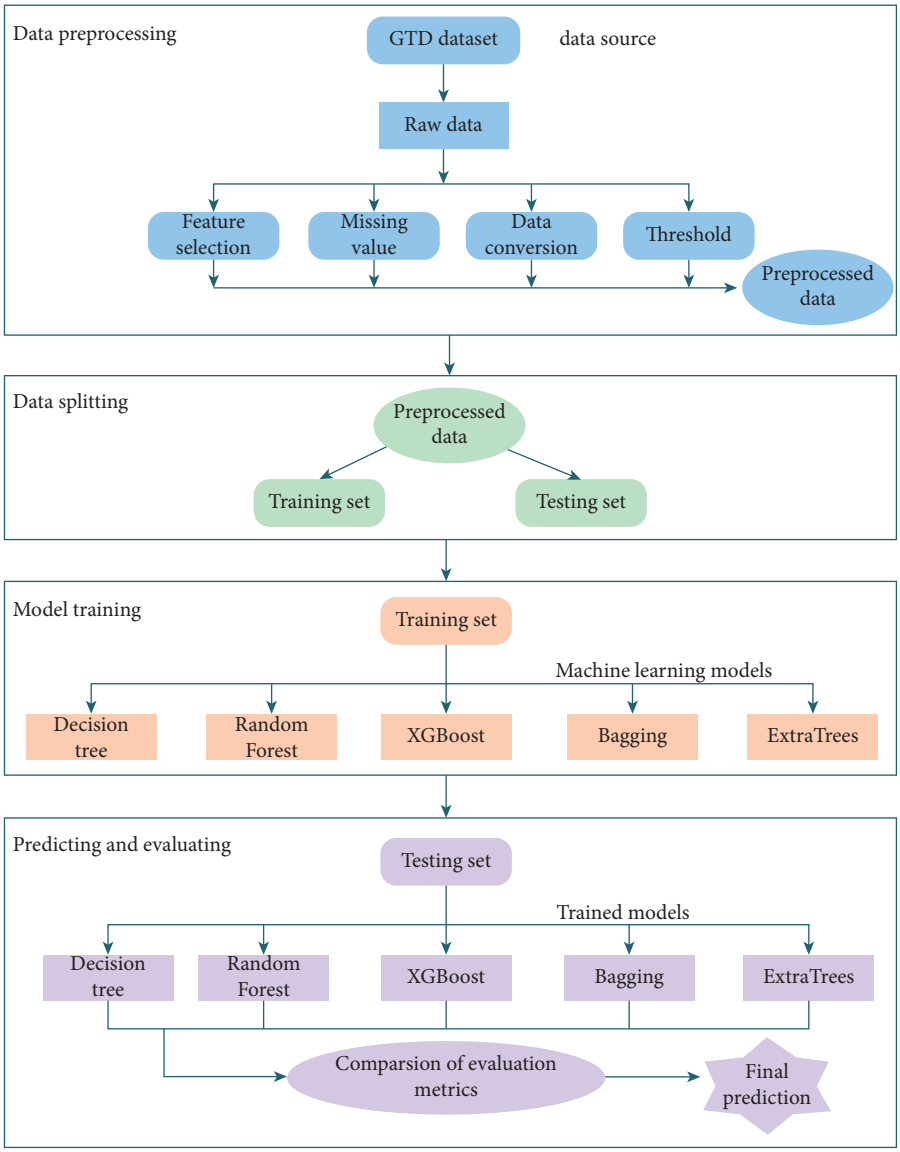

Figure 4: The framework for classifying and predicting terrorist organizations.

Table 2: Confusion matrix.

Actual category	Predicted category	
	True	False
True	True positive	False negative
False	False positive	True negative

In the evaluation of multiclass classification, P, R, and $F1$ scores were used to evaluate each category. The overall evaluation of all multiple categories is usually performed using accuracy and macroaverage. Macroaverage first calculates the statistical indicators precision, recall, and $F1$ for each category separately and then calculates the arithmetic average for all categories.

4. Experiment Methodology and Result Analysis

Quantitative analysis and modeling prediction of global terrorist attacks were performed in Python 3.6, running on a platform with an Intel Core i7 processor and 24.00 GB DDR RAM. We utilized the Python libraries pandas-0.25.2, numpy-1.17.2, xgboost-1.0.0, and scikit-learn-0.21.3 [31].

For visualization of the analysis results, we used seaborn-0.9.0 and matplotlib- 3.1.1 in Python.

4.1. Data Structure Analysis. The data primarily contained the following attributes of information: GTD serial number, date, event description information, time, location, attack description information, weapon information, target information, victim information, casualty information, and action results. There were many fields under each type of information to enrich the data. Each terrorist attack was stored as a record (i.e., a row) of 137 attributes such as country, year, number of deaths and injuries, and use of weapons. Among them, there were 46 attributes with a completeness of more than 70%.

4.2. Data Preprocessing. In the dataset, the average number of attacks by all terrorist organizations was 28. However, 3,430 terrorist organizations (91% of all terrorist organizations) launched fewer than 28 terrorist attacks, and 2,600 terrorist organizations (73% of all organizations) launched fewer than five terrorist attacks. These 2,600 terrorist organizations launched 4,038 terrorist attacks, which accounted for only 4% of the identified terrorist attacks (i.e., attacks by identified terrorist organizations). If all terrorist organizations were predicted, too many categories and low-sample categories may cause unfavorable training interference noise. Therefore, to make the experiment closer to reality and the trained model more effective, samples with fewer than five terrorist attacks were removed in this study.

Some attributes are unrelated to the prediction of terrorist organizations. Training on these attributes would not only increase the required training time but also render the training results unreasonable or impractical; therefore, data preprocessing operations are essential. At this stage, the GTD dataset was processed through data cleaning, feature engineering, and data normalization.

4.2.1. Data Cleaning. Data cleaning aims to reduce the dimensions of the GTD dataset by detecting and deleting irrelevant or redundant attributes and case records.

First, attribute fields that contained descriptive text or too many missing values (the missing threshold was set to 30%) were removed. Second, missing values in specific attribute fields were filled with the numerical value corresponding to "unknown" according to the data description rules provided by the GTD. Third, some attribute fields were converted into numerical values to facilitate later processing. For example, the "related" attribute field provides the "eventid" of other terrorist attacks' related to this terrorist attack in text format, and we convert it to the count of related terrorist attacks. The number of event records after these three steps was reduced to 98,909. Fourth, after deleting the records of terrorist attacks with fewer than five terrorist attacks, we filtered the remaining records of terrorist attacks according to five conditions (i.e., ≥5 times, ≥50 times, ≥100 times, ≥500 times, ≥1000 times). Eventually, the number of records in the experimental dataset was reduced to 94871 after the data cleaning process.

4.2.2. Feature Engineering. We tended to retain the objective attributes of the terrorist attacks in the GTD and ignored some subjective judgment criteria and some text columns for interpretation and apparently irrelevant attributes, such as crit1-3, country_txt, region_txt, and eventID. Therefore, 45 possible related attributes were left for analysis. Further selections were then made. First, 34 numerical data (int, float) were selected without special processing. Then, the target/victim nationality (natlty1-3) was transformed into int numerical type and was selected. In this way, 37 candidate feature attributes were filtered out. For these 37 properties, it is difficult to determine the feature attributes that should be retained or removed because the remaining feature attributes after data cleaning are correlated in some way. After considering some solution strategies, we used the feature selection function (i.e., SelectKBest) in sklearn to make the selection, and very few adjustments were made.

Based on the above strategies, to simplify the model and improve the prediction accuracy, we selected 36 features for the experiment, including iyear, imonth, iday, extended, country, region, successful attack, suicide attack, attack type1-3, target type1-3, target subtype1-3, target nationality (natlty1-3), weapon type1-4, weapon subtype1-4, property, ishostkid, ransom, related, INT_IDEO, INT_LOG, INT_MISC, and INT_ANY. We used the ExtraTrees classifier to build a forest to rank the importance of the 36 feature attributes, as shown in Figure 5. It may be observed that the three attributes, namely, the country and region where the terrorist attack occurred and the target nationality were the most critical attributes for predicting terrorist organizations.

Thus, the GTD was transformed into a new dataset with a scale of $94871 * 37$ after data preprocessing. Among them, "gname" is the target attribute for prediction, and the remaining 36 attributes are the explanatory features for prediction.

4.3. Data Splitting. In machine learning, the sample dataset is usually partitioned into a testing set and a training set in proportion. Because the classification of the target feature attributes in the dataset is usually unevenly distributed, the training and testing sets are divided according to the proportion of the target features in the sample dataset, such that the proportion of the data in each category of the training set and the testing set is consistent with the proportion of the sample dataset, thereby reducing the misleading predictions of the trained models. The following two methods are generally used in data splitting.

4.3.1. Hold Out. Directly partition the data (Data) into two mutually exclusive sets, one of which is used as the training set (Training) and the other as the testing set (Testing), $\text{Data} = \text{Training} \cup \text{Test}$, $\text{Training} \cap \text{Testing} = \varnothing$. When dividing the data into a training set and a testing set, the data

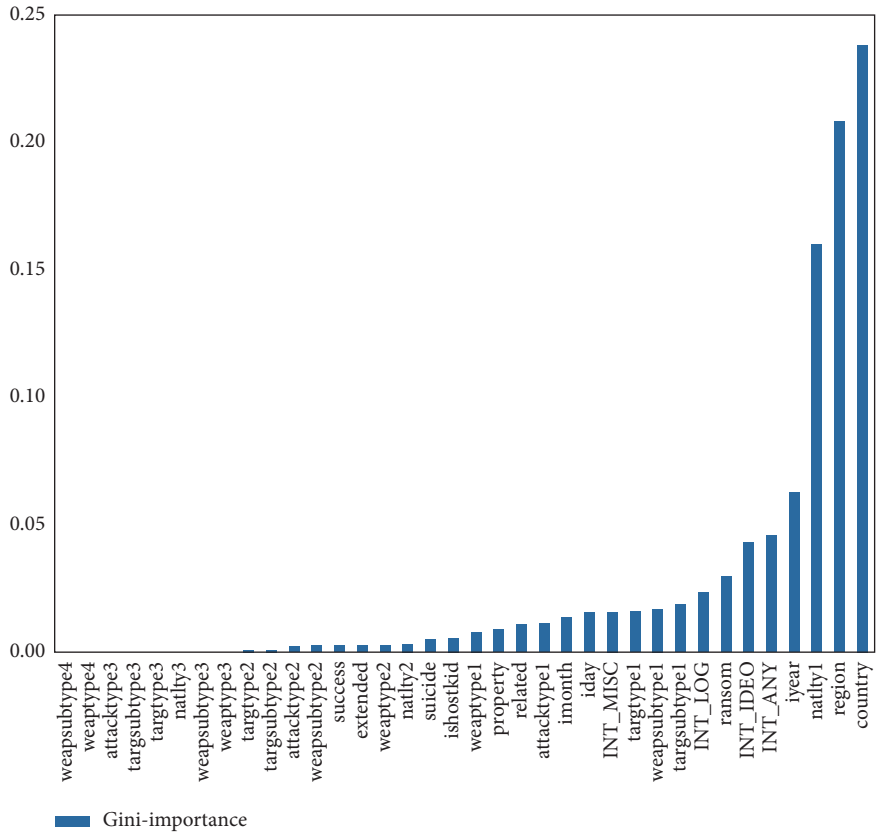

Figure 5: Importance ranking of 36 feature attributes.

consistency must be maintained as much as possible to avoid affecting the final result from deviations in the data division process.

4.3.2. K-Fold Cross Validation. Divide the data into k mutually exclusive subsets of similar size, namely, Data $= D_1 \cup D_2 \cup \cdots \cup D_k$, $D_i \cap D_j = \varnothing\,(i \neq j)$. Each time, the union of $k-1$ subsets was used as the training set, and the remaining subset was used as the testing set. Thus, k combinations of the testing and training sets were obtained. The k-fold cross-validation method is a common method used to alleviate the problem of data imbalance.

In this study, the hold-out method was used to split the sample dataset into 90% and 10% portions as the training and testing sets, respectively. The training set was used to build the model, and the testing set was used to test and evaluate the effectiveness of the model. In addition, we used the 10-fold cross-validation method to evaluate the models further and compared the results with those of the hold-out method to verify the stability of the model.

4.4. Model Construction and Evaluation Analysis

4.4.1. Classifier Modeling. In this study, five mainstream classifiers, including DT, Bagging, RF, ET, and XGBoost, were used to classify and predict terrorist attack organizations that perpetrated specific attacks. We first optimized these models and then evaluated and compared the performances of these models.

Hyperparameters [32] in the machine learning models can be manually set and continuously optimized by trial-and-error minimization. Table 3 briefly introduces the major parameters used in the ensemble learning models.

We optimized these hyperparameters by performing grid searches or random searches for manual tuning of the model and then comparing the accuracy and selecting the optimal parameter value for model performance. After tuning, we obtained the optimal parameters of these five models as follows: The DT uses the CART algorithm by default, and the parameter random_state is 42. Bagging uses the KNN classifier as the base classifier, and the random_state, max_features, and max_samples are 30, 0.5, and 0.5, respectively. The RF parameters n_estimators, max_features, and random_state are 500, "sqrt," and 42, respectively. The ET parameters n_estimators, min_samples_split, and random_state are 10, 2, and 12, respectively. The XGBoost parameter n_estimators is 300. All other parameters of the above models were set to default values.

We performed the following two experiments using the hold-out method and the 10-fold cross-validation method and compared the results to verify the performance and stability of the models.

(1) Hold-out method: the new data obtained after data preprocessing is split into a training set containing

Table 3: Description of parameters in ensemble learning models.

Parameter	Description
*n*_estimators	Number of base classifiers
random_state	Number of seeds of random number generator
max_features	Maximum number of features involved in judgment when splitting nodes
max_samples	Maximum number of samples
min_samples_split	Minimum number of samples required for splitting

90% data and a testing set containing 10% data using the "hold-out" method, and the above five classification models are trained and tested, respectively, to obtain the corresponding accuracy, precision, recall and *F*1.

(2) 10-fold cross-validation method: owing to the imbalance of terrorist organization type data, the 10-fold cross-validation method is a common method to alleviate the problem of data imbalance. For instance, some terrorist organizations performed a large number of terrorist attacks, while other terrorist organizations performed a small number of attacks. Therefore, the five algorithms were trained ten times through the 10-fold cross-validation method, the output of each training result was retained, and the average value of the ten output results was calculated to obtain the cross-validation accuracy.

4.4.2. Result Evaluation and Analysis. To comprehensively reflect the global prediction performance, we performed predictions in each of the five ranges of terror attack frequency, and the experimental results of the model performance experiment in terms of accuracy, precision, recall, and *F*1 indicators are shown in Table 4 and Figure 6.

It may be clearly seen from Figure 6 that the 19 terrorist organizations with the highest terror attack frequency (≥1000) had the highest prediction accuracy for all algorithms, and the prediction accuracy of terrorist organizations decreased as the frequency of terror attacks decreased. For 936 terrorist organizations with more than five terrorist attacks, the prediction accuracy was reduced to approximately 0.85. This indicates that as the number of predicted target classes for the multiclassification problem increased substantially, the algorithm's ability to discriminate some features decreased, and it tended to confuse the target classes with high similarity. From the comparison of the five algorithms, it may be seen that the XGBoost model performs best in the case of the number of terrorist attacks (≥1000, ≥500), and the RF model performs best in the case of the number of terrorist attacks (≥100, ≥50, ≥5).

To explore the analysis results visually, we focused on terrorist organizations with frequent attacks. Here, we analyze in detail the experimental results of 32 terrorist organizations with no less than 500 terrorist attacks frequency. The experimental results showed that the prediction accuracy of the XGBoost model reached 97.1634%, with a precision of 95.7246%, and the comprehensive evaluation index *F*1 was 95.0011%. Random Forests was closely followed, with an accuracy of 96.8216%. We observed that the three models achieved the highest accuracy among all five models, that is, XGBoost, Random Forests, and ExtraTrees. These models are all ensemble learning algorithms based on the tree model; therefore, it is concluded that the tree model's ensemble learning classifier was optimal for this research.

To compare the performance of the models before and after alleviating the imbalance of the types of terrorist organization data, the five models were trained and tested using the hold-out method (without alleviating data imbalance) and the 10-fold cross-validation method (with alleviating data imbalance), respectively. The accuracies of the 32 terrorist organizations are shown in Table 5 and Figure 7.

Although the precision of the 10-fold cross-validation method was slightly lower, the accuracy, recall, and *F*1 were almost the same. The 10-fold cross-validation method can alleviate data imbalance and avoid misleading models; thus, the models reflect more realistic and effective results. The model metric difference between the two methods is very small, which also shows that the constructed prediction models are relatively stable and accurate.

5. Visual Exploration and Discussion

To observe the effect of the models visually, we used the confusion matrix of 32 terrorist organizations to show the prediction results. In confusion matrices, we can observe the number of correct and incorrect predictions and the results of incorrect predictions (e.g., terrorist organization A is predicted to be terrorist organization B).

Because the XGBoost model outperforms all other models (as shown in Table 5), it is meaningful to explore and analyze the effect of the XGBoost model. The visualization of the confusion matrix for the prediction results of the XGBoost model is shown in Figure 8. Other models can also be visually analyzed in the same manner.

There are 5,852 records (i.e., terrorist attacks) in the test data after data splitting. We draw the confusion matrix of the XGBoost model for 32 terrorist organizations with more than 500 terrorist attacks, as shown in Figure 8. The rows of the confusion matrix represent the actual terrorist organizations in the test data, and the columns of the confusion matrix represent the predicted terrorist organizations in the test data. Therefore, the diagonal values in the confusion matrix represent the number of terrorist attacks in which the corresponding terrorist organization was predicted correctly, while the values outside the diagonal represent the number of corresponding terrorist organizations that were incorrectly predicted as being responsible for attacks. For a model that achieves 100% accurate prediction, the confusion

Table 4: Comparison of 5 algorithms for predicting terrorist groups with different attack frequencies.

Summary of data records	Number of terrorist attacks (range) / Number of terrorist organizations / Total number of terrorist attacks	≥1000 19 50200	≥500 32 58520	≥100 122 78107	≥50 210 84339	≥5 936 94871
Accuracy	Decision trees	0.982669	0.958647	0.878377	0.854636	0.796164
	Bagging	0.960757	0.931989	0.833312	0.799858	0.740620
	Random forests	0.983068	0.968216	**0.904494**	**0.881195**	**0.835687**
	ExtraTrees	0.979283	0.959501	0.886698	0.860327	0.803225
	XGBoost	**0.983466**	**0.971634**	0.853924	0.791439	0.698567
Precision	Decision trees	0.976950	0.928242	0.787521	0.745648	0.478754
	Bagging	0.945956	0.932152	0.771253	0.685609	0.347113
	Random forests	**0.979232**	**0.957727**	**0.847559**	**0.817384**	**0.520747**
	ExtraTrees	0.973327	0.942159	0.811242	0.761273	0.476816
	XGBoost	0.978406	0.957246	0.752235	0.523490	0.126893
Recall	Decision trees	**0.976073**	0.929554	**0.786034**	0.737920	**0.512161**
	Bagging	0.940090	0.858106	0.605144	0.523339	0.304769
	Random forests	0.974743	0.934140	0.785265	**0.739619**	0.511993
	ExtraTrees	0.970025	0.926234	0.761510	0.708550	0.469377
	XGBoost	0.976059	**0.944904**	0.746525	0.537858	0.138689
*F*1 score	Decision trees	0.976497	0.928603	0.784185	0.736470	0.481310
	Bagging	0.941057	0.875151	0.633059	0.551191	0.305712
	Random forests	0.976587	0.942883	**0.805488**	**0.754597**	**0.502975**
	ExtraTrees	0.971609	0.932798	0.779096	0.723834	0.459432
	XGBoost	**0.977118**	**0.950011**	0.745925	0.523800	0.130349

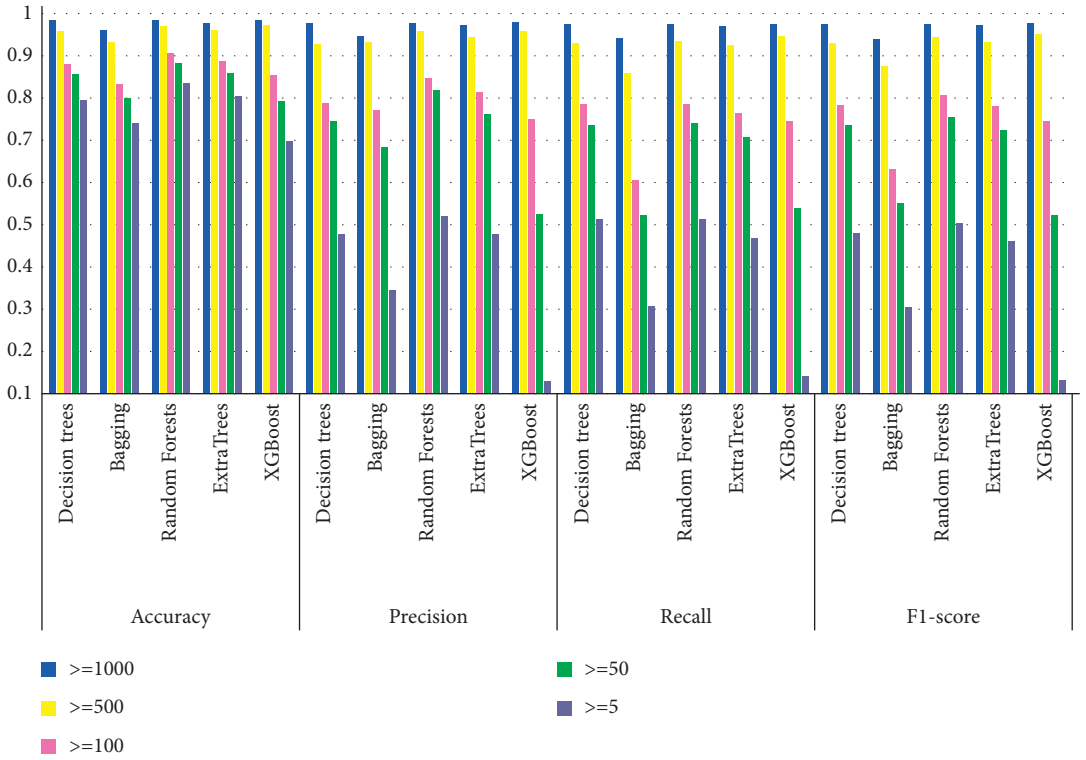

Figure 6: Overall evaluation and comparison of 5 models.

matrix is a diagonal matrix. Otherwise, nonzero numbers in the off-diagonal positions of the confusion matrix represent the number and result in incorrect predictions. For instance, the value 1 in row 2, column 9 indicates that a terrorist attack by the African National Congress (South Africa) was incorrectly predicted as the Corsican National Liberation Front (FLNC).

From Figure 8, it can be observed that the confusion matrix was almost diagonally symmetric. The terrorist organizations with the most predicted errors were the National

Table 5: Comparison of 10-fold cross-validation and hold-out methods (terror attack frequency ≥ 500).

Metrics		Data split verification method			
		Hold-out method	10-fold cross-validation method		
			Mean	Max	Min
Accuracy	Decision trees	0.958647	0.956026	0.964387	0.949943
	Bagging	0.931989	0.933528	0.936764	0.927569
	Random forests	0.968216	0.965974	0.968744	0.962647
	ExtraTrees	0.959501	0.959406	0.962400	0.956011
	XGBoost	**0.971634**	**0.967778**	0.970828	0.963216
Precision	Decision trees	0.928242	0.929096	0.943624	0.918073
	Bagging	0.932152	0.927242	0.938100	0.911614
	Random forests	**0.957727**	**0.955594**	0.963753	0.947774
	ExtraTrees	0.942159	0.941508	0.945737	0.935197
	XGBoost	0.957246	0.952817	0.956800	0.943972
Recall	Decision trees	0.929554	0.931822	0.944081	0.923149
	Bagging	0.858106	0.862504	0.871233	0.854989
	Random forests	0.934140	0.934952	0.941422	0.929029
	ExtraTrees	0.926234	0.928944	0.934792	0.923533
	XGBoost	**0.944904**	**0.942287**	0.950476	0.933696
*F*1 score	Decision trees	0.928603	0.930063	0.943702	0.920590
	Bagging	0.875151	0.875903	0.886239	0.866793
	Random forests	0.942883	0.942658	0.949366	0.935658
	ExtraTrees	0.932798	0.933608	0.937390	0.927087
	XGBoost	**0.950011**	**0.946542**	0.953123	0.937474

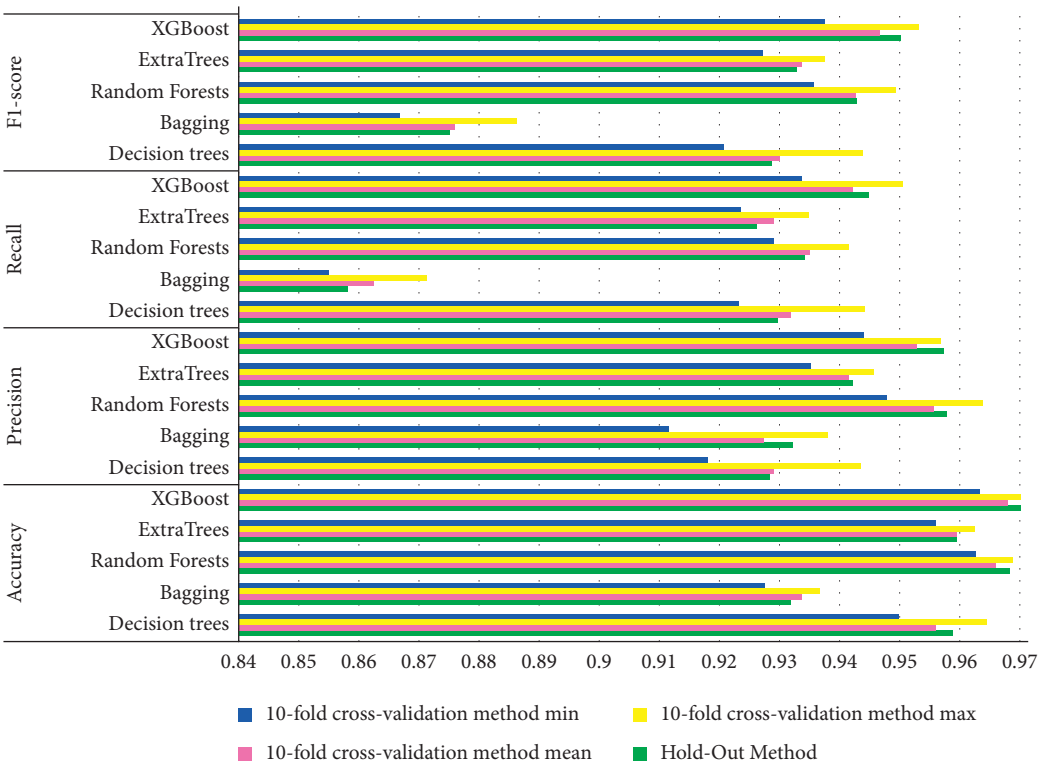

Figure 7: Comparison of model metrics using two data split verification methods.

Liberation Army of Colombia (ELN) in line 11 from the bottom and the Revolutionary Armed Forces of Colombia (FARC) in line 7 from the bottom. 27 attacks by the ELN were erroneously predicted as FARC, and 25 attacks of FARC were erroneously predicted as ELNs. These values are close to diagonally symmetric. The reason is that these two terrorist organizations have similar features, so the prediction model can easily be confused with them. Similarly, the organizations FARC and M-19 (Movement of April 19) were also similar, with 9 and 14 prediction errors,

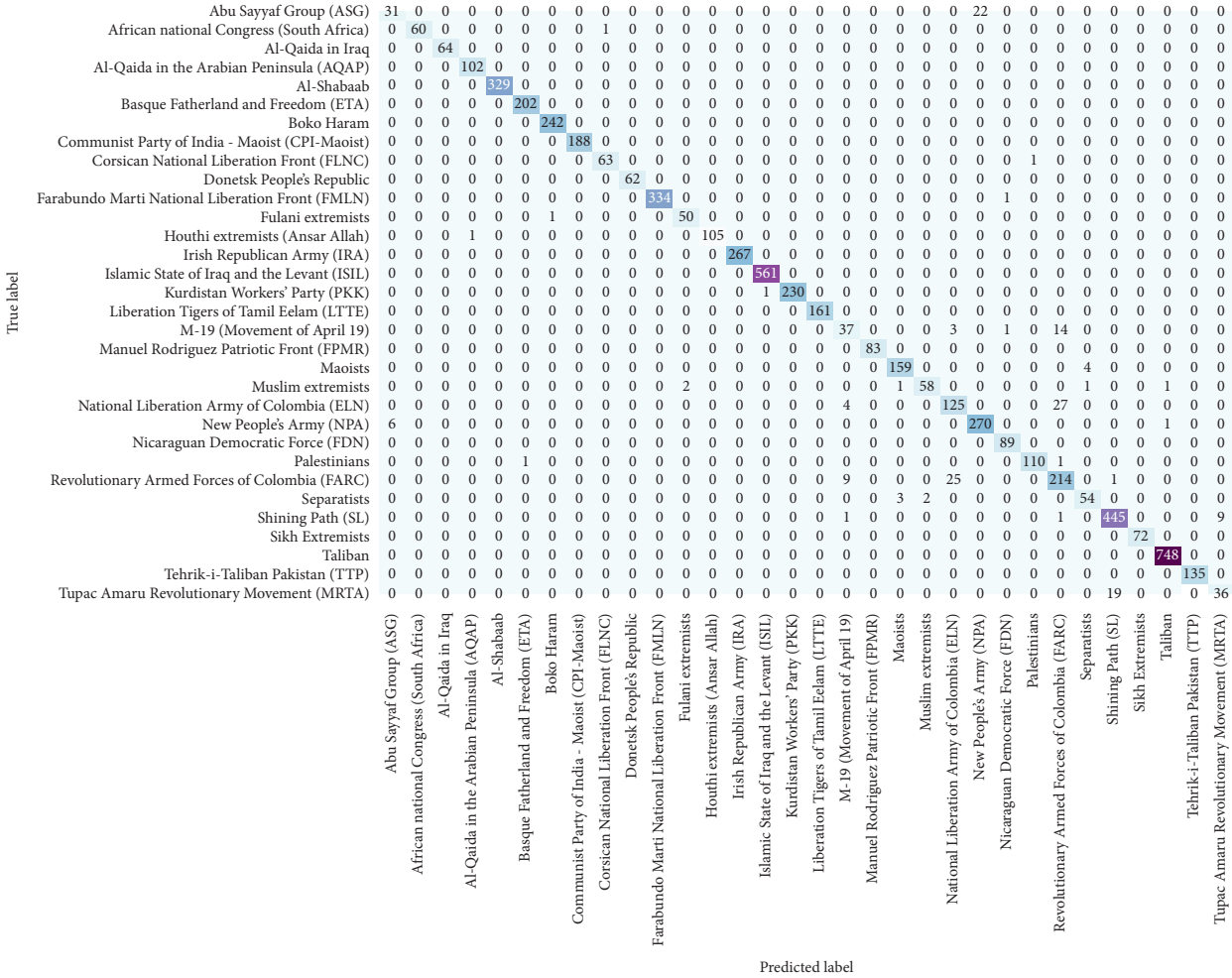

Figure 8: Confusion matrix of 32 terrorist organization prediction of the XGBoost model.

respectively. Therefore, it may be observed that FARC was more easily incorrectly predicted than other organizations, and its prediction precision and recall were correspondingly lower.

6. Conclusion

In this study, through a quantitative analysis of the data in the GTD, ensemble machine learning has been used to construct five multiclass classification models for the prediction of terrorist organizations that perpetrated terrorist attacks.

First, according to the frequency of terrorist organization attacks, the terrorist organizations were analyzed, and the characteristics and trends of 32 terrorist organizations with more than 500 terrorist attacks were described in detail. Then, for the prediction of terrorist organizations in terrorist attacks, 36 feature attributes were selected based on the feature selection strategy, and five classifiers, including decision tree, bagging, random forest, extra tree, and XGBoost, were constructed to predict terrorist organizations. The performance and stability of the five models were evaluated using hold-out and 10-fold cross-validation methods, respectively. Our models predicted 32 terrorist organizations for high-frequency activities in terrorist attacks. Finally, the experimental results showed that the five models achieved good performance and stability. XGBoost and the random forest classifier achieved the best prediction accuracies of 97.15% and 97.03%, respectively. We further visualized and analyzed the prediction results of the XGBoost model using the confusion matrix. Moreover, the method can be extended to the prediction of a broader range of terrorist organizations. Considering the number of terrorist organization classifications based on the frequency of attacks, the classification prediction accuracy of the random forest algorithms was consistently excellent. When the number of terrorist organizations was small (e.g., dozens), XGBoost exhibited the best prediction accuracy, and the performance of random forest was close to that of XGBoost.

The prediction model presented herein can macroscopically predict the terrorist organizations of global terrorist attacks, excavate the relevant factors of terrorist attacks, and provide decision support for the prevention and control of antiterrorism organizations and related countries. With further improvement in the performance and accuracy of machine learning algorithms, we believe that these

technologies can help security departments find better algorithmic models and appropriate datasets to improve the accuracy of predictions related to terrorist attacks. However, considering the local sparsity of terrorist attacks and their versatility in planning and execution, even with the continuous progress of machine learning, research on large-scale monitoring and prediction algorithms is nonetheless expected to be challenging.

Data Availability

The dataset used in this work is derived from the data on terrorist attacks from 1970 to 2017 in the Global Terrorism Database (GTD) [2]. The GTD dataset is built by the United States Anti-Terrorism Research Consortium (START) and the University of Maryland and is publicly available at https://www.start.umd.edu/gtd/.

Conflicts of Interest

The author declares no conflicts of interest.

Acknowledgments

This work was supported in part by the Philosophy and Social Science Planning Project of Shanghai, under Grant no. 2019BGL028.

References

[1] J. M. Poland, *Understanding Terrorism: Groups, Strategies, and Responses*, Prentice-Hall, Englewood Cliffs, 1988.

[2] G. LaFree and L. Dugan, "Introducing the global terrorism database," *Terrorism and Political Violence*, vol. 19, no. 2, pp. 181–204, 2007.

[3] F. Ding, Q. Ge, D. Jiang, J. Fu, and M. Hao, "Understanding the dynamics of terrorism events with multiple-discipline datasets and machine learning approach," *PloS One*, vol. 12, no. 6, Article ID e0179057, 2017.

[4] Y.-L. Chuang, N. Ben-Asher, and M. R. D'Orsogna, "Local alliances and rivalries shape near-repeat terror activity of al-Qaeda, ISIS, and insurgents," *Proceedings of the National Academy of Sciences*, vol. 116, no. 42, pp. 20898–20903, 2019.

[5] V. B. Petroff, J. H. Bond, D. H. Bond, and D. H. Bond, "Using hidden Markov models to predict terror before it hits (again)," in *Handbook of Computational Approaches to Counterterrorism*, pp. 163–180, Springer, New York, NY, USA, 2013.

[6] F. Gohar, W. H. Butt, and U. Qamar, "Terrorist group prediction using data classification," in *Proceedings of the International Conference on Artificial Intelligence and Pattern Recognition*, pp. 199–208, Kuala Lumpur, Malaysia, 2014.

[7] G. M. Tolan, O. S. Soliman, and O. S. Soliman, "An experimental study of classification algorithms for terrorism prediction," *International Journal of Knowledge Engineering-IACSIT*, vol. 1, no. 2, pp. 107–112, 2015.

[8] X. Meng, L. Nie, and J. Song, "Big data-based prediction of terrorist attacks," *Computers & Electrical Engineering*, vol. 77, pp. 120–127, 2019.

[9] B. Bu, Z. Pi, and L. Wang, "Support vector machine for classification of terrorist attacks based on intelligent tuned harmony search," *Ekoloji*, vol. 28, no. 107, pp. 153–164, 2019.

[10] Z. Li, D. Sun, B. Li, Z. Li, and A. Li, "Terrorist group behavior prediction by wavelet transform-based pattern recognition," *Discrete Dynamics in Nature and Society*, vol. 2018, Article ID 5676712, 2018.

[11] X. Hu, F. Lai, G. Chen, R. Zou, and Q. Feng, "Quantitative research on global terrorist attacks and terrorist attack classification," *Sustainability*, vol. 11, no. 5, p. 1487, 2019.

[12] G. M. Campedelli, M. Bartulovic, and K. M. Carley, "Learning future terrorist targets through temporal meta-graphs," *Scientific Reports*, vol. 11, no. 1, pp. 8533–8615, 2021.

[13] G. M. Campedelli, *On Meta-Networks, Deep Learning, Time and Jihadism*, Università Cattolica del Sacro Cuore, XXXII ciclo, a.a. 2018/19, Milanohttp://hdl.handle.net/10280/70552.

[14] G. M. Campedelli, I. Cruickshank, and K. M. Carley, "A complex networks approach to find latent clusters of terrorist groups," *Applied Network Science*, vol. 4, no. 1, pp. 1–22, 2019.

[15] B. A. Desmarais and S. J. Cranmer, "Forecasting the locational dynamics of transnational terrorism: a network analytic approach," *Security Informatics*, vol. 2, no. 1, pp. 1–12, 2013.

[16] W. Guo, K. Gleditsch, and A. Wilson, "Retool AI to forecast and limit wars," *Nature*, vol. 562, no. 7727, pp. 331–333, 2018.

[17] S. B. Kotsiantis, I. Zaharakis, and P. Pintelas, "Supervised machine learning: a review of classification techniques," *Emerging artificial intelligence applications in computer engineering*, vol. 160, pp. 3–24, 2007.

[18] S. R. Safavian and D. Landgrebe, "A survey of decision tree classifier methodology," *IEEE transactions on systems, man, and cybernetics*, vol. 21, no. 3, pp. 660–674, 1991.

[19] P. E. Utgoff, N. C. Berkman, and J. A. Clouse, "Decision tree induction based on efficient tree restructuring," *Machine Learning*, vol. 29, no. 1, pp. 5–44, 1997.

[20] R. J. Lewis, "An introduction to classification and regression tree (CART) analysis," in *Annual Meeting of the Society for Academic Emergency Medicine in San Francisco*, vol 14, AddisonWesley Educational, California, USA, 2000.

[21] L. Rutkowski, M. Jaworski, L. Pietruczuk, and P. Duda, "The CART decision tree for mining data streams," *Information Sciences*, vol. 266, pp. 1–15, 2014.

[22] T. G. Dietterich, "Ensemble learning," *The handbook of brain theory and neural networks*, vol. 2, pp. 110–125, 2002.

[23] L. Breiman, "Bagging predictors," *Machine Learning*, vol. 24, no. 2, pp. 123–140, 1996.

[24] M. Galar, A. Fernandez, E. Barrenechea, and H. Sola, "A review on ensembles for the class imbalance problem: bagging-, boosting-, and hybrid-based approaches," *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 42, no. 4, pp. 463–484, 2011.

[25] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.

[26] P. Geurts, D. Ernst, and L. Wehenkel, "Extremely randomized trees," *Machine Learning*, vol. 63, no. 1, pp. 3–42, 2006.

[27] L. Wehenkel, D. Ernst, and P. Geurts, "Ensembles of extremely randomized trees and some generic applications," in *Proceedings of Robust Methods for Power System State Estimation and Load Forecasting*, Paris, 2006.

[28] T. Chen and C. Guestrin, "Xgboost: a scalable tree boosting system," in *Proceedings of the 22nd ACM sigkdd international conference on knowledge discovery and data mining*, pp. 785–794, San Francisco, CA, USA, August 2016.

[29] T. Chen, T. He, and M. Benesty, *Xgboost: Extreme Gradient Boosting*, pp. 1–4, 2015, R package version 0.4-2.

[30] Y. Kong and M. Jing, "Research of the classification method based on confusion matrixes and ensemble learning," *Computer Engineering & Science*, vol. 34, no. 6, p. 111, 2012.

[31] F. Pedregosa, G. Varoquaux, A. Gramfort, and V. Michel, "Scikit-learn: machine learning in python," *The Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

[32] J. Bergstra and Y. Bengio, "Random search for hyper-parameter optimization," *Journal of Machine Learning Research*, vol. 13, no. Feb, pp. 281–305, 2012.