Preprint Manuscript: Bridgelall, R. (2023). Identifying Factors Associated with Terrorist Attack Locations by Data Mining and Machine Learning. *International Social Science Journal*. (Online) DOI:10.1111/issj.12414.

Identifying Factors Associated with Terrorist Attack Locations by Data Mining and Machine Learning

Raj Bridgelall, Ph.D.
Associate Professor, Transportation, Logistics & Finance, College of Business
North Dakota State University
PO Box 6050, Fargo ND 58108-6050
Email: raj@bridgelall.com
ORCiD: 0000-0003-3743-6652

Compliance with Ethical Standards
Funding: The author did not receive support from any organization for the submitted work.
Declarations: The author has no relevant financial or non-financial interests to disclose.
Data Availability: The sources of the datasets used in this study are cited within the manuscript, and they are all publicly available.

Identifying Factors Associated with Terrorist Attack Locations by Data Mining and Machine Learning

Abstract

While studies typically investigate the socio-economic factors of perpetrators to comprehend terrorism motivations, there was less emphasis placed on factors related to terrorist attack locations. Addressing this knowledge gap, this study conducts a multivariate analysis to determine attributes that are more associated with terrorist attacked locations than others. To tackle the challenge of identifying pertinent attributes, the methodology merges a global terrorism database with relevant socio-economic attributes from the literature. The workflow then trains 11 machine learning models on the combined dataset. Among the 75 attributes assessed, 10 improved the predictability of targeted locations, with population and public transportation infrastructure being key factors. After optimizing hyperparameters, a multi-layer perceptron—a type of artificial neural network—exhibited superior predictive performance, achieving an AUC score of 89.3%, classification accuracy of 88.1%, and a harmonically balanced precision and recall score of 87.3%. In contrast, support vector machines demonstrated the poorest performance. The study also revealed that race, age, gender, marital status, income level, and home values did not improve predictive performance. The machine learning workflow developed can aid policymakers in quantifying risks and making objective decisions regarding resource allocation to safeguard public health.

Keywords: Counterterrorism; Data Fusion; Exploratory Spatial Data Analysis; Feature Relevance Scoring; Multivariate Analysis; Population Demographics; Predictive Models; Transportation Security

Conflict of Interest: none

1 Introduction

Although research frequently explores the socio-economic factors of perpetrators to understand terrorism motivations, there has been less attention to factors associated with terrorist attack locations. Mass shootings, particularly in the United States, often prompt questions about the reasons behind terrorists targeting specific locations (Metzl and MacLeish 2015). One might assume that perpetrators consistently choose large cities with numerous targets such as public transportation, resulting in severe consequences. However, due to terrorism's adaptive nature and multifaceted motives (Bridgelall 2022), targeted locations can vary widely. Applied intelligence to identify attributes more closely related to attacked locations can assist policymakers in prioritizing risk mitigation and countermeasure resources for locations at elevated risk. It is essential to note that a statistical association between attributes and targeted locations does not necessarily imply latent or causal relationships.

The **objective** of this research was to leverage data mining (DM) and machine learning (ML) methods to identify statistical associations of various attributes with attacked locations. The approach was to mine the literature on terrorism to identify plausible attributes, combine various datasets that contain those attributes, and then train predictive ML models on the combined dataset. Developers of the Global Terrorism Database (GTD)™ define terrorism as “the threatened or actual use of illegal force and violence by a non-state actor to attain a political, economic, religious, or social goal through fear, coercion, or intimidation.” (START 2020). To control for attribute heteroscedasticity due to political variations across large regions, the case study focused on data about populated places within the United States.

A few studies investigated attributes that are associated with terrorist attack locations. For example, Nussio et al. (2021) provided evidence that terrorists seek locations that maximize

attention getting (Nussio, Bohmelt and Bove 2021). Blomberg et al. (2013) found that terrorism is mostly unrelated to the economic conditions of attacked locations (Blomberg, Fernholz and Levin 2013). Korotayev et al. (2019) found that terrorist attack intensity decreased in more developed locations that promoted education (Korotayev, Vaskin and Tsirel 2019). More terrorism occurred in locations that restricted economic freedom (Gassebner and Luechinger 2011), and locations in early stage democracy (Iheonu, et al. 2021). Terrorists also tend to target locations with major international gateways such as ports because of the potential for massive disruptions in the supply chain (Young and Gordon 2020). Terrorists also targeted transportation systems to disrupt mobility, especially in landlocked countries (Mlepo 2022).

Only recently has data science methods such as ML begun to creep into studies involving psychology and human behavior (Grimm and Jacobucci 2021). Nevertheless, Canhoto (2021) found that the key challenge is a lack of datasets that can leverage ML techniques to help in the fight against terrorism (Canhoto 2021). In related work, Huamaní et al. (2020) applied tree-based ML models to predict attack frequency within 12 regions of the world and found that Random Forest (RF) provided the best performance with a predictive accuracy of 89.5% (Huamaní, Alicia and Roman-Gonzalez 2020). Ding et al. (2017) applied RF, Artificial Neural Networks (ANN), and Support Vector Machines (SVM) to a database of global terrorism and predicted attack regions with an accuracy of 96.6% (Ding, et al. 2017). Hao et al. (2019) found that RF coupled with a geographical information system (GIS) provided good prediction of location within the Indochina Peninsula (Hao, et al. 2019). Li et al. (2021) used k-means clustering, an unsupervised ML approach, to classify terrorist activities by region, attack type, target type, and weapon type (Li, et al. 2021). Overall, there was a scarcity of studies that utilized data science methods to identify characteristics or attributes of terrorist attack locations.

Contributions of this work are as follows:

1. A hybrid DM and ML workflow that policy makers can use to generalize and replicate the analysis (Section 3).
2. A list of relevant datasets and their sources to support further analysis by the research or counterterrorism community (Section 3.2).
3. A detailed chronicle of the dataset preparation to substantially reduce the time that analysts would spend when using the same data (Section 3.2). The impact of this contribution is based on the finding that data scientists spend an average of 60% of their time cleaning and organizing data (Ilyas and Chu 2019).

It is difficult to know which ML method works best for a given application (James, et al. 2013). Therefore, this work evaluated 11 of the most mature types of ML models on the uniquely fused dataset to compare their predictive performance.

The organization of the rest of this paper is as follows: Section 2 describes the selection of attributes and datasets, model types, attribute engineering, methods to compare the ML performance, and methods to rank the relevance of attributes selected. Section 3 presents the specific models used, their results, and attribute ranking. Section 4 discusses the significance of the results. Section 5 concludes the work and outlines future work that will apply the workflow to data from other regions of the world for comparison.

2 Methods and Data

Figure 1 shows the workflow developed to identify a heterogenous set of attributes that are associated with terrorist attack locations. The ML models represented data on historical attack locations. Hence, evaluating the predictive performance of a ML model explains the relative degree to which each attribute is associated with attack locations. The next subsections describe

the main procedures within the workflow, which the author coded in software. Those procedures include attribute initialization, data fusion, attribute engineering, ML, predictive performance evaluation, and attribute relevance scoring.

The workflow shows that there were iterations between the data preparation and ML layers in two optimization loops to facilitate data cleaning and ML model adjustments. The first loop back to the data fusion procedures adjusted the data extraction and data cleaning actions based on the results of the data merging. For example, some of the merge keys had different spellings or formats that needed correcting. The second loop was from the ML to the data fusion procedures. The adjustments included aspects of the data cleaning, attribute engineering, attribute selection, and hyperparameter tuning to maximize the predictive performance.

The subsections that follow provide more details on all the workflow procedures shown in Figure 1. Subsection 3.1 discusses the EDA and DM to identify an initial set of relevant attributes. Subsection 3.2 describes the datasets used. Subsection 3.3 discusses the attribute engineering to maximize information gain. Subsection 3.4 describes the different ML methods evaluated. Subsection 3.5 describes the metric to evaluate ML predictive performance. Finally, subsection 3.6 describes the methods used to score the relevance of attributes.

2.1 Attribute Initialization

The selection of an initial set of attributes to test for relevance in improving the predictive performance of each ML method was based on two strategies. The first strategy was to apply EDA on the GTD to derive spatial or physical attributes that were associated with frequent attack targets. The second strategy was to review the existing body of knowledge on terrorism to empirically identify common underlying socio-economic factors of locations where terrorism tends to occur.

Figure 2 shows the EDA result for frequency of attacked targets in the United States. The numbers and percentages shown next to each bar indicates the amount and proportion of the data that each accounted for. The split color bars show the relative proportion of failed attacks for that category. The distribution shown in Figure 2 revealed that businesses, private citizens, and general governments were the top three categories of targets, which accounted for more than half (51.2%) of the total attacks. The color coding in the chart shows the relative proportion of attacks that succeeded in each target category. Further EDA revealed that the most frequently attacked places in the business targets were banks and their branches. The next dominant category was "private citizens & property" which suggested that characteristics of both the environment and the people at a location were influential attributes. Table 1 summarizes the *spatial* (location sensitive) attributes selected from the EDA process and the rationale for their selection. Table 2 summarizes the demographic attributes considered and the hypotheses for their selection.

2.2 Dataset Selection

Informed by the attribute initialization, Table 3 summarizes the datasets used and lists their dimensions and sources. The extracted FTA, APTA-F, and APTA-V datasets are available from the Federal Transit Administration (FTA), American Public Transportation Association (APTA) facility database, and the APTA vehicle database, respectively. The population and TIGER® geographic feature datasets are available from the Bureau of Transportation Statistics (BTS) and the U.S. Census Bureau (USCB), respectively. At the time of acquisition, the GTD contained records of terrorism events from 1970 to 2018. The separate GTD dataset for 1993 was a reconstruction of previously missing data for that year. The author licensed the "People" data set from Pareto Software, which is a compilation of publicly available U.S. demographic datasets

from various government organizations and census. The “Worship” dataset is available from the Association of Statisticians of American Religious Bodies (ASARB).

The “Parks” dataset is available from the U.S. Geological Survey (USGS). The next nine datasets from “Banks” through “Airports” are available from the Homeland Infrastructure Foundation-Level Data (HIFLD). The last two datasets on energy profile are available from the National Renewable Energy Laboratory (NREL). The subsections that follow chronicle the data extraction, cleaning, merging, and imputation applied to each dataset.

2.2.1 Public Transit Datasets

Table 7 (appendix) summarizes the operations applied to each of the three public transit datasets before merging them.

FTA: The “Facility Type” subset of the “Facilities and Stations” dataset from the “2019 Annual Data Tables” contained 2,779 agency records with 31 attributes describing the number and types of facilities including maintenance stations, fueling facilities, administrative offices, passenger stations and passenger stops. The records also listed the number of passenger vehicles that operated to provide maximum system capacity. The selected attributes were the number of passenger stations aggregated from several types such as elevated, at-grade, underground, grade-separated, ferryboat, and bus transfer stations. The initial merge attempt revealed records with duplicate city names due to inconsistent or incorrect spelling. Examples include “Swan Quarter” versus “Swans Quarter” and “Sault Sainte Marie” or “Sault Ste. Marie” versus “Sault Ste Marie” which a cleaning strategy of sorting after each merge revealed as orphaned neighbors.

APTA-F: The “Station Data” subset of the “Infrastructure Database” contained 501 agency records with 43 attributes describing their facility types and amenities such as security cameras,

concessions, and elevators. The selected attributes were the number of passenger stations and passenger stops, aggregated by agency.

APTA-V: The "Active Vehicles" subset of the "US Fleet" database contained 6,936 records of each passenger vehicle type. There were 36 attributes describing the vehicle characteristics such as fuel type, size, make, model, year built, handicap accessibility equipment, and purchase price. The selected attributes were the number of vehicles of any type that each agency operated.

Merge-T: A pivot table operation for each dataset aggregated the attributes by all agencies operating in unique U.S. cities. The dataset was missing a Federal Information Processing Standard (FIPS) code to uniquely identify each city. Therefore, the merge procedure created a unique key by concatenating the city name and the state abbreviation. The merge resulted in discrepancies in the number of stations and vehicles for some cities. The resolution was to take the maximum value. The merge operation helped to reduce the number of missing values for stations and vehicles for 46 and 2 cities, respectively.

Transit: The attribute that represented accessibility (Table 2) was the number of passenger transit stops in each populated place. Values were missing for 2,006 cities after merging the transit datasets. The strategy to impute the number of stops was a two-layer estimate based on the number of vehicles and the number of stations. The first imputation layer filled missing values for stops by computing the average number of stops per operating vehicle from cities where those valuables were present. The average was 8.5, so the estimated number of stops for places where the number of vehicles was available was the ceiling of 8.5 times the number of vehicles. The second imputation layer filled the remaining missing values for stops by computing the average number of stops per station where those values were present. The average was 137.24 so the imputed number of stops for places where the number of stations was present was

the ceiling of 137.24 times the number present. Finally, the procedure dropped 380 records that had no data. Hence, the final transit dataset contained 1,911 U.S. cities with the number of stops present or estimated.

2.2.2 Populated Places

The BTS dataset described in Table 3 contained the 2010 census population and the average elevation for 38,186 places in the United States. However, the population was missing for 11,125 cities. There were also 241 cities with the same name in a state. The duplicate city and state names prevented the creation of a unique key for merging with the combined transit dataset. Hence, combining records of the duplicate cities facilitated the merge. The merge and sort cleaning strategy revealed 92 records in the combined transit dataset with city names that did not match those in the BTS dataset for several reasons. Table 8 (appendix) summarizes the reasons for those mismatches into six categories of data cleaning. After manually correcting the city names, the repeated merge-sort cycles revealed that the BTS was missing 37 cities in the combined transit dataset. A manual Internet search created those missing records.

The attributes selected from the BTS dataset to merge with the combined transit dataset were the average elevation, the population from the 2010 census, the county name, and the FIPS code. Table 9 (appendix) summarizes the merging and cleaning operations that created the fused dataset in the data preparation layer of the workflow. There were no missing values for the 1,911 locations after the final merge-sort iteration.

2.2.3 Global Terrorism Database

The next merge iteration added the target attribute (dependent variable), which was a binary flag indicating whether a city ever experienced a terrorist attack. Table 10 (appendix) chronicles the operations used to complete the merge. The first step was to extract records for attacks in U.S.

cities. The second step was to append the reconstructed records for the year 1993. This time, the first merge attempt revealed that there were 26 cities with mismatched spelling. The spelling errors encountered fell within the categories listed in Table 8 (appendix), for example, “De Witt” versus “DeWitt” versus “Dewitt” and “Saint Charles” versus “St. Charles” and others. Finally, a pivot table operation summarized the number of attacks within each city and added the unique merge key. The merge revealed that terrorist attacks never occurred in 1,563 (81%) cities in the transit dataset. The iterative process of the data fusion continued with the addition of attributes selected from the other datasets as described in the next subsections.

2.2.4 City Demographics

The author purchased a comprehensive “People” dataset that contained the demographics of more than 108,000 U.S. cities (SimpleMaps 2020). The curator combined information from the United States Geological Survey and the 2018 American Community Survey completed by the United States Census Bureau. This analysis used the cleaned dataset released on November 18, 2020. Once again, the next merge iteration revealed 41 mismatched city names and some mismatched FIPS code in the previously fused transit and GTD datasets. After correcting the 41 records, the procedure selected the attributes listed in Table 11 (appendix) from the “People” dataset. The table indicates the number of missing values per attribute after the merge. There was a total of 61 attributes across the 23 categories listed in the table. That is, some of the categories listed are aggregates of their sub-category stratifications. For example, “Education” was available as separate attributes for the population proportion that completed high school, college, and graduate school.

2.2.5 County Datasets

The next 14 data merge operations used the FIPS code that was either available or spatially derived for each of the remaining datasets listed in in Table 3. It was possible to derive the FIPS code for datasets that did not include them only if they included a geospatial coordinate that represented the centroid a location. For those cases, a GIS spatial join procedure with the TIGER/Line® shapefile for U.S. counties (USCB 2019) merged the FIPS code.

2.3 Model Selection

Several textbooks describe the theories of operation for each of the model type selected. Géron (2017) discusses the decision tree (DT), random forest (RF), AdaBoost (AB), logistic regression (LR), support vector machine (SVM), stochastic gradient descent (SGD), and multi-layer perceptron (MLP) models and demonstrates their operation using the Python programming language (Géron 2017). James et al. (2013) explains naïve Bayes (NB), k-nearest-neighbors (kNN), and tree-based *bagging* and *boosting* methods (James, et al. 2013). Hastie et al. (2016) discusses how to train models to improve their *generalized* performance (Hastie, Tibshirani and Friedman 2016). Murphy (2012) shares insights about model operation from a probabilistic perspective (Murphy 2012). James et al. (2013) describes how the process of *k-fold cross validation* prevents the training process from *overfitting* or *underfitting* to improve their predictive performance (James, et al. 2013).

2.4 Attribute Engineering

The next subsections describe three methods of attribute engineering that improved the ML predictive performance. They were attribute selection based on the data fusion, attribute transformation, and attribute normalization.

2.4.1 Attribute Transformation and Normalization

Some of the ML methods require or assume that the attributes have a normal distribution. However, the distribution of several attributes such as population and the number of stops for public transit had a heavy right skew. Conversely, some of the other distributions such as "Race" skewed heavily to the left. A shifted natural logarithm, LN(1 + x) where x is the attribute value, reduced the right skew and a squared transformation reduced the left skew. The transformed attributes improved the ML predictive performance over the non-transformed versions of those attributes.

The performance of some algorithms such as gradient descent improves when all attributes have a comparable range (Géron 2017). Therefore, the workflow normalized all attributes $\boldsymbol{x}$, including the transformed ones, to the [0, 1].

2.4.2 Attribute Selection

Some of the attributes such as the number of places of worship, banks, and colleges did not improve the ML predictive performance until normalized by other attributes. In particular, the spatial density or inverse spatial density improved the ML predictive performance for some attributes, but not all of them. The inverse spatial density improved predictive performance for places of worship, college campuses, banks, government buildings, Department of Defense (DoD) facilities, Environmental Protection Agency (EPA) facilities, parks, housing units (HU), and vehicle miles traveled (VMT). Conversely, the inverse spatial density decreased predictive performance for police facilities, courts, Fortune 500 campuses, hospitals, and airports.

Table 4 summarizes the new attributes evaluated and the number of missing values. Attributes with more than 50% missing values decreased the ML predictive performance so the procedure dropped those. Missing values imputed from the mean value did not change the ML

predictive performance for attributes with a small proportion of missing values. Other attributes, not listed, such as the land-to-water area ratio and the elevation-to-population ratio also decreased ML predictive performance, so the workflow eliminated those in the optimization loop.

2.5 Predictive Performance

The literature offers several metrics to measure the predictive performance of classification models. The most common are classification accuracy (CA), precision, and recall. CA is simply the ratio of the number of correctly classified instances to the total number of instances. The main disadvantage of CA is that it can be misleading when the dataset is highly imbalanced in class representation (Krawczyk 2016). That is, a no-skill classifier that predicts the dominant class every time can appear to perform better.

To understand all performance metrics, it is helpful to think of them in terms of a radio receiver that must distinguish a signal pulse from a noise pulse. *Precision* is a measure of how precise the classifier is at predicting positive instances. That is, precision = TP/(TP + FP) where TP and FP represent the number of true positives and false positives, respectively. Hence, a low detection threshold might increase precision by detecting weak signals. However, the receiver may also misclassify noise pulses as signal pulses (false positive) if the threshold is too low, thus reducing precision. *Recall* is a measure of the classifier's ability to correctly predict all the positive samples. That is, recall = TP/(TP + FN), where FN represents the number of false negatives, which are the missed positives. Hence, a threshold that is too far above the noise level may reduce FP but miss true signals, thus reducing the recall rate. The F1 score is a balanced measure of precision and recall as their harmonic mean where

$$F1 = \frac{TP}{TP + \left(\frac{FP + FN}{2}\right)} \quad (1)$$

The CA, precision, recall, and F1 scores can produce misleading results when the data is highly imbalanced. A more suitable metric is the area under the curve (AUC) of the receiver operating characteristic (ROC). The curve is a plot of the TP rate against the FP rate, both as a function of the classifiers output of class membership probability (Fawcett 2006). A no-skill classifier will produce an AUC score near 0.5 whereas the best performing models will have AUC scores approaching 1.0.

2.6 Relevance Scoring

The relevance of an attribute in helping to distinguish among target classes is proportional to the amount of separation in the distributions of that attribute for each class (Agresti 2018). For example, the distribution of neck length in a sample population of animals in a large zoo will have little overlap between those of the adult giraffes and those of the adult elephants. Two effective measures of the statistical difference between distributions are the Pearson's χ^2 (Chi-Squared) statistic and the Cohen's *d* effect size. The χ^2 statistic is associated with a p-value from the chi-squared distribution that determines if there is a statistically significant difference between two distributions (Agresti 2018). The χ^2 statistic is

$$\chi^2 = \sum_{i=1}^{N} \frac{(u_i - v_i)^2}{v_i} \quad (2)$$

where u_i and v_i are values for an attribute in the first and second classes, respectively. A statistic value of zero means that the distributions are identical, and larger values indicate greater differences.

The Cohen's *d* effect size (Sawilowsky 2009) is a simpler measure that uses the mean and standard deviation of the attribute within each class. The statistic is

$$d = \frac{|\mu_1 - \mu_2|}{\sqrt{\frac{\sigma_1^2 + \sigma_2^2}{2}}} \quad (3)$$

where μ_1 and μ_2 are the mean values of the distributions of the first and second classes, respectively. The standard deviations of the two distributions are σ_1 and σ_2. Sawilowsky (2009) established a "rule of thumb" that values greater than 0.8 and lower than 0.2 indicate large and slight differences, respectively.

3 Results

The next two subsections discuss the results of the predictive performance measures for the 11 different models and the relevance scoring for those attributes that improved predictive performance.

3.1 Predictive Performance

Table 5 lists the 11 models and the five predictive performance metrics calculated for each in the order of their AUC score. There were differences in the predictive performance of each ML model. Seven of the models provided both an AUC score and classification accuracy above 0.86. MLP and XGB were the best-performing models. However, MLP had the edge with respect to the F1 score. The MLP model exhibited the highest AUC score (0.893), classification accuracy (0.881), and F1 score (0.873). The MLP also demonstrated strong precision (0.872) and recall (0.881) metrics. This indicated that the MLP model had the best overall predictive performance. Both the MLP and XGB models can adeptly deal with attribute nonlinearities and noisy datasets. However, MLP has the disadvantage of a longer training time, which can become computationally expensive as the size and dimensionality of the data grows. As expected, the no-skill classifier had the lowest AUC score, and the CA is equal to the class imbalance, which reflected that there were no attacks in 81.8% of the locations.

The SVM model displayed the lowest AUC score (0.677) and classification accuracy (0.578). However, it achieved relatively high precision (0.792), while its recall score remained low (0.578). An intuitive explanation for why SVM performed poorly is that the model seeks clean hyperplanes in multidimensional datasets to separate classes. Therefore, SVM will perform poorly when the dataset is noisy and the distribution of attribute values within each class has large overlap. Figure 3 illustrates that there was a high degree of overlap among the distributions of the relevant attributes. The height of the bars indicates the relative frequency of the values on the horizontal axis, within a nominal fixed bin size. The bars split the frequency of each value bin by the target class. That is, class 1 and class 0 indicate attacks and no attacks, respectively, for each value bin. The line curves are Gaussian distributions with the same means and standard deviations as those of the attribute within each class. The high overlap increases the difficulty of classification.

3.2 Relevance Scoring

Table 6 lists only those attributes that improved the predictive performance of the top models, in the order of their effect size. That is, the listed attributes improved the performance of the MLP and XGB models. The χ^2 and $\boldsymbol{d}$ columns of the table lists the Chi-squared statistics and the effect sizes, respectively. The average effect size was 0.93, which the literature considers to be large. The correlation between the effect size and the Chi-squared statistic was 0.903, which indicates good agreement in the ranking of attribute relevance. The “Attacked” column indicates the direction of the attribute value range that was associated with attacked locations, where “H” and “L” indicates high and low values, respectively. For example, the probability that a location would experience an attack increased with its population size, the number of access points

(stops) to public transportation, and the vehicle traffic density. The discussion section below further interprets the relevance scorings and their implications.

Figure 3 shows the distribution of the two highest ranking attributes by effect size. The curves serve only to help visualize the amount of separation between the distributions. Figure 4 are visualizations of the effect size for two of the attributes that did not improve the ML classification performance. Their effect sizes were two to three times smaller than the average effect size for those attributes that did contribute towards improving the ML classification performance.

4 Discussion

The results of this work suggest that ML models can represent the location of terrorist attacks with reasonable accuracy if there are enough relevant attributes. The workflow demonstrated that a specific ordering of procedures with iterative cycles to manipulate the data and tune the models is necessary to maximize performance. That is, the procedures must appropriately clean, transform, normalize, and impute missing data for the models to be effective. Furthermore, the process showed that the art of attribute engineering to identify relevant attributes and to derive new attributes by combining irrelevant attributes is an important requirement. The interpretation of attributes that reduced classification accuracy is that they manifest as noise in the dataset and do not contribute to any discernable structure or pattern.

Only 10 of the 75 attributes evaluated helped to improve the predictive performance of the best models. A key finding was that population size, the number of access points for public transit, and traffic density were more significantly associated with attacked locations. This finding validates the intuition that larger, busier, and more accessible cities would be more attractive to perpetrators than sparsely populated places. This finding supports the recent results

from Nussio et al. (2021) (Nussio, Bohmelt and Bove 2021). Another finding that was consistent with historic attacks in the United States is that attacked locations had a higher density of hospitals, places of worship, and banks.

An interesting finding was that attacked locations tended to have a larger proportion of the population with education levels beyond college. An unexpected finding was that places where people spoke limited English had a higher probability of attack than places with large non-white or migrant populations. A possible explanation is that people in United States territories where attacks occurred likely speak less English than places in the continental United States with large non-white or migrant populations where no attacks occurred.

There were several surprising results for the U.S. case study that were contrary to findings that compared different countries. Demographic attributes that did not improve the predictive performance of the ML models included race, age, gender, marital status, family size, income level, home ownership levels, and home values. A possible explanation is that the literature generally examined the socio-economic situations of perpetrators or their homeland rather than those of locations that experienced terrorist attacks. This is an important distinction that points to the value of studying characteristics of the perpetrators and their homelands, as well as attributes of the targeted locations.

Limitations of using DM and ML methods to represent attack locations based on significant attributes are the potential for a high number of false positives and a high number of false negatives. Both types of errors can result in either unnecessary expenditure to bolster security or complacency that could increase attack risks. Furthermore, the workflow is difficult to automate for ongoing predictions based on new data. For example, a data scientist must identify relevant attributes and tune the hyperparameters of each model to maximize their predictive performance.

The data preparation procedures of the workflow were the most time-consuming steps. Examples of hyperparameters are the type of regularization, the number estimators in ensemble models, the learning rate in boosting models, and the network configuration of the MLP. Incidentally, the MLP tuned in this analysis performed best with 30 nodes in a single hidden layer. In general, approaches that use some form of artificial intelligence require multidisciplinary experts with both data science and counterterrorism knowledge to select relevant attributes, to optimize each model, to select the best models, to interpret the results, and to synthesize appropriate recommendations. Therefore, risk managers should understand these limitations when considering the workflow for use in their organizations or cities.

5 Conclusion

Building artificial intelligence through data mining and machine learning can effectively identify patterns in data to inform strategies and policies to protect public health. The data mining and machine learning workflow developed in this study evaluated 61 demographic and 14 physical attributes for their relevancy in representing terrorist attack locations based on the model's predictive performance. Of the 75 total attributes evaluated, the workflow identified 10 that improved the predictive performance of the top ML models. The most relevant attribute was the population of the location, which was either a city, town, village, or municipality.

The next most important attributes were related to transportation. Those attributes were the number of stops for public transit and the traffic density, measured as the accumulated travel distance of vehicles per unit area of land in the county of the populated place. The other important attributes were physical or spatial features, which were the number of hospitals, the number of law enforcement facilities, the density of banks, and the density of places of worship. Of the 61 demographic attributes evaluated, only two were relevant; they were the proportion of

the population that spoke limited English and the proportion of the population that had greater than college-level education. Other demographic attributes such as the distribution of race, age, gender, marital status, family size, income level, home ownership levels, and home values did not improve the predictive performance of the classification models.

Seven of the ML models provided good predictive performance with an AUC score better than 0.87. Multi-layer perceptron (MLP), which is a type of artificial neural network, provided the best overall predictive performance with an AUC score of 0.893, a classification accuracy of 88.1%, and an F1 score of 87.3%.

The implications of the findings of this research are that analysts can leverage the potential of artificial intelligence and machine learning in informing counter-terrorism strategies and policies, helping to allocate resources efficiently and protect public health. The importance of the population attribute suggests that policymakers should pay particular attention to densely populated areas, such as cities, towns, villages, or municipalities, when allocating resources to counter terrorism threats. The significance of transportation-related attributes suggests that analysts should consider public transit infrastructure and traffic density when developing counter-terrorism strategies. The relevance of features such as hospitals, law enforcement facilities, banks, and places of worship, implies that these locations may require heightened security measures.

The best configuration for the MLP on the fused dataset was 30 nodes in a single hidden layer. This result suggests that deep learning with more complex artificial neural networks architectures can potentially improve the predictive performance when using much larger datasets with even more attributes. Hence, future work will investigate the performance of

various deep learning models on datasets from other regions of the world for comparison of the results.

References

Agresti, Alan. 2018. *Statistical Methods for the Social Sciences.* 5th. Boston, Massachusetts: Pearson.

APTA-I. 2018. *Infrastructure Database.* December 30. Accessed January 1, 2021. https://www.apta.com/research-technical-resources/transit-statistics/transit-infrastructure-database/.

APTA-V. 2020. *Public Transportation Vehicle Database.* December 30. Accessed January 10, 2021. https://www.apta.com/research-technical-resources/transit-statistics/vehicle-database/.

ASARB. 2020. *US Religion Census 2010.* Association of Statisticians of American Religious Bodies (ASARB). November 30. Accessed November 30, 2020. http://www.usreligioncensus.org/.

Blomberg, S. Brock, Ricardo Fernholz, and John-Clark Levin. 2013. "Terrorism and the Invisible Hook." *Southern Economic Journal* 79 (4): 849-863. doi:10.4284/0038-4038-2012.290.

Bridgelall, Raj. 2022. "An Application of Natural Language Processing to Classify What Terrorists Say They Want." *Social Sciences* 11 (1): 23. doi:10.3390/socsci11010023.

BTS. 2019. *Populated Places.* Bureau of Transportation Statistics (BTS). May 14. Accessed June 25, 2020. https://data-usdot.opendata.arcgis.com/datasets/populated-places.

Canhoto, Ana Isabel. 2021. "Leveraging machine learning in the global fight against money laundering and terrorism financing: An affordances perspective." *Journal of Business Research* 131: 441-452. doi:10.1016/J.JBUSRES.2020.10.012.

Ding, Fangyu, Quansheng Ge, Dong Jiang, Jingying Fu, and Mengmeng Hao. 2017. "Understanding the Dynamics of Terrorism Events with Multiple-Discipline Datasets and Machine Learning Approach." *PLoS ONE* 12 (6): e0179057.

Fawcett, Tom. 2006. "An introduction to ROC analysis." *Pattern Recognition Letters* (Elsevier) 27 (8): 861-874. http://people.inf.elte.hu/kiss/11dwhdm/roc.pdf.

FDIC. 2021. *Institutions & Locations.* Federal Deposit Insurance Corporation (FDIC). January 7. Accessed January 9, 2021. https://www.fdic.gov/bank/statistical/.

FTA. 2020. *The National Transit Database (NTD).* Federal Transit Administration (FTA). April 6. Accessed June 24, 2020. https://www.transit.dot.gov/ntd.

Gassebner, Martin, and Simon Luechinger. 2011. "Lock, stock, and barrel: a comprehensive assessment of the determinants of terror." *Public Choice* 149 (3): 235-261. doi:10.1007/S11127-011-9873-0.

Géron, Aurélien. 2017. *Hands-On Machine Learning with Scikit-Learn and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems.* 2nd. Sebastopol, California: O'Reilly Media.

Grimm, Kevin J., and Ross Jacobucci. 2021. "Reliable Trees: Reliability Informed Recursive Partitioning for Psychological Data." *Multivariate Behavioral Research* 56 (4): 595-607. doi:10.1080/00273171.2020.1751028.

Hao, Mengmeng, Dong Jiang, Fangyu Ding, Jingying Fu, and Shuai Chen. 2019. "Simulating Spatio-Temporal Patterns of Terrorism Incidents on the Indochina Peninsula with GIS and the Random Forest Method." *ISPRS International Journal of Geo-Information* 8 (3): 133.

Hastie, Trevor, Robert Tibshirani, and Jerome Friedman. 2016. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction.* 2nd. New York, New York: Springer.

HIFLD. 2017. *Homeland Infrastructure Foundation-Level Data (HIFLD).* U.S. Department of Homeland Security. Accessed January 9, 2021. https://hifld-geoplatform.opendata.arcgis.com/.

Hohl, Alexander, Moongi Choi, Aggie J. Yellow Horse, Richard M. Medina, Neng Wan, and Ming Wen. 2022. "Spatial Distribution of Hateful Tweets Against Asians and Asian Americans During the COVID-19 Pandemic, November 2019 to May 2020." *American journal of public health* 112 (4): 646-649. doi:10.2105/AJPH.2021.306653 .

Huamaní, Enrique Lee, Alva Mantari Alicia, and Avid Roman-Gonzalez. 2020. "Machine Learning Techniques to Visualize and Predict Terrorist Attacks Worldwide using the Global Terrorism Database." *International Journal of Advanced Computer Science and Applications* 11 (4). doi:10.14569/IJACSA.2020.0110474.

Iheonu, Chimere O., Simplice A. Asongu, Shedrach A. Agbutun, and Innocent A. Ifelunini. 2021. "Democracy and terrorism in Africa." *International Social Science Journal.* doi:10.1111/issj.12317.

Ilyas, Ihab F., and Xu Chu. 2019. *Data Cleaning.* New York, NY: Association for Computing Machinery and Morgan & Claypool Publishers.

James, Gareth, Daniela Witten, Trevor Hastie, and Robert Tibshirani. 2013. *An Introduction to Statistical Learning with Applications in R.* Vol. 112. New York: Springer. doi:10.1007/978-1-4614-7138-7.

Korotayev, Andrey, Ilya Vaskin, and Sergey Tsirel. 2019. "Economic Growth, Education, and Terrorism: A Re-Analysis." *Terrorism and Political Violence* 1-24. doi:10.1080/09546553.2018.1559835.

Krawczyk, Bartosz. 2016. "Learning from Imbalanced Data: Open Challenges and Future Directions." *Progress in Artificial Intelligence* 5 (4): 221-232. doi:10.1007/s13748-016-0094-0.

Li, Zhongbei, Xiangchun Li, Chen Dong, Fanfan Guo, Fan Zhang, and Qi Zhang. 2021. "Quantitative analysis of global terrorist attacks based on the global terrorism database." *Sustainability* 13 (14): 7598. doi:10.3390/su13147598.

Ma, Ookie, Ricardo P Cardoso de Oliveira, Evan Rosenlieb, and Megan H Day. 2019. *Sector-Specific Methodologies for Subnational Energy Modeling.* NREL/TP-7A40-72748, Golden, Colorado: National Renewable Energy Laboratory (NREL). doi:10.2172/1506626.

Metzl, Jonathan M., and Kenneth T. MacLeish. 2015. "Mental illness, mass shootings, and the politics of American firearms." *American journal of public health* 105 (2): 240-249. doi:10.2105/AJPH.2014.302242 .

Mlepo, Andrew Thomas. 2022. "Attacks on road-freight transporters: a threat to trade participation for landlocked countries in Southern Africa." *Journal of Transportation Security* 1-18. doi:10.1007/s12198-021-00242-6.

Murphy, Kevin P. 2012. *Machine Learning : A Probabilistic Perspective.* Cambridge, Massachusetts: The MIT Press.

Nussio, Enzo, Tobias Bohmelt, and Vincenzo Bove. 2021. "Do Terrorists Get the Attention They Want? Comparing Effects of Terrorism across Europe." doi:10.1093/poq/nfab046.

Sawilowsky, Shlomo S. 2009. "New Effect Size Rules of Thumb." *Journal of Modern Applied Statistical Methods* 8 (2): 26. doi:10.22237/JMASM/1257035100.

SimpleMaps. 2020. *Simple Maps.* Vers. 1.71. LLC Bright Market. November 18. Accessed November 18, 2020. https://simplemaps.com/data/us-cities.

START. 2020. National Consortium for the Study of Terrorism and Responses to Terrorism (START). April 13. Accessed April 13, 2020. https://www.start.umd.edu/data-tools/global-terrorism-database-gtd.

USCB. 2019. *TIGER/Line Shapefiles Technical Documentation.* Washington, D.C.: United States Census Bureau (USCB), 138. https://www2.census.gov/geo/tiger/TIGER2019/COUNTY/.
USGS. 2018. *PAD-US Data Download.* United States Geological Survey (USGS). Accessed January 9, 2021. doi:10.5066/P955KPLE.
Young, Richard R., and Gary A. Gordon. 2020. "Intermodal maritime supply chains: assessing factors for resiliency and security." *Journal of Transportation Security* 13 (3): 231-244. doi:10.1007/s12198-020-00224-0.

6 Figures and Tables

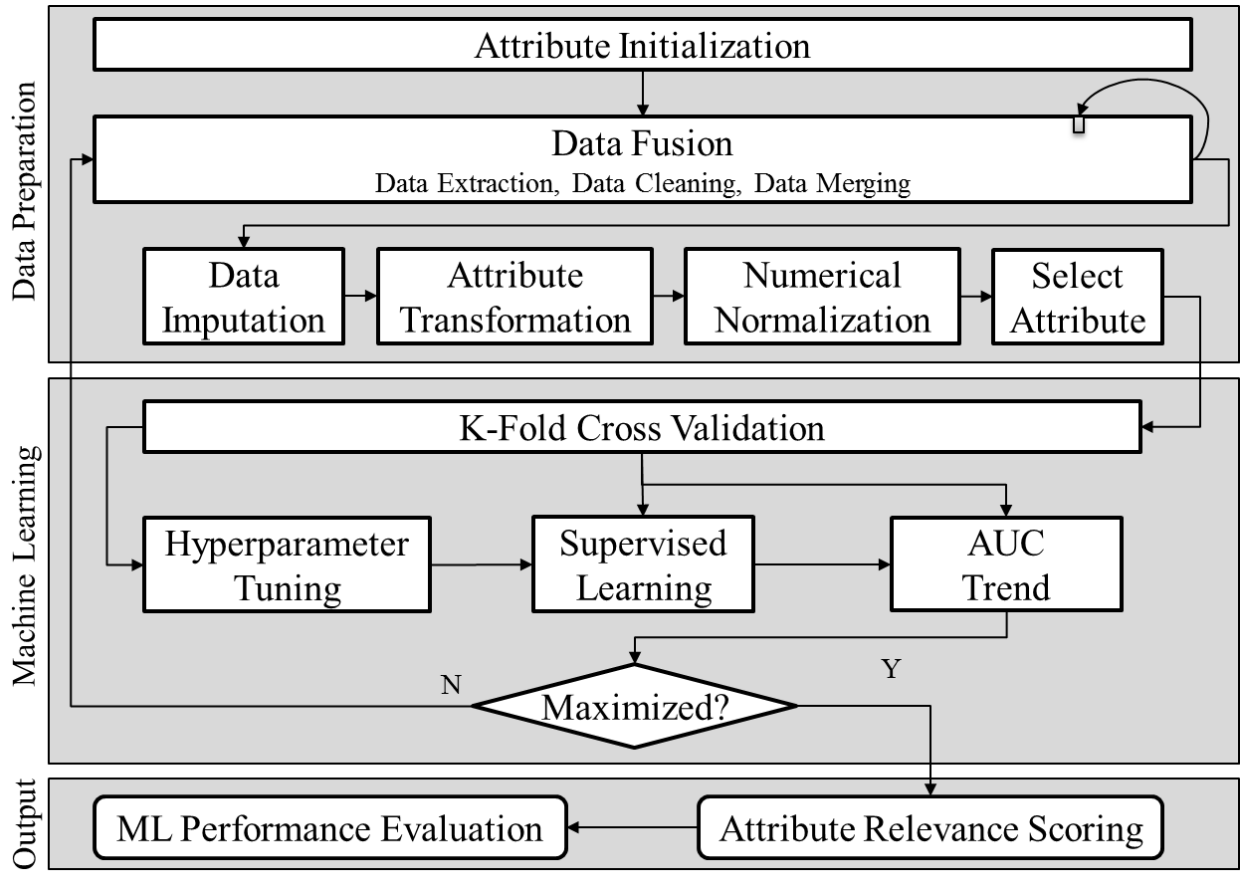

Figure 1: Workflow for data fusion, machine learning, and attribute selection.

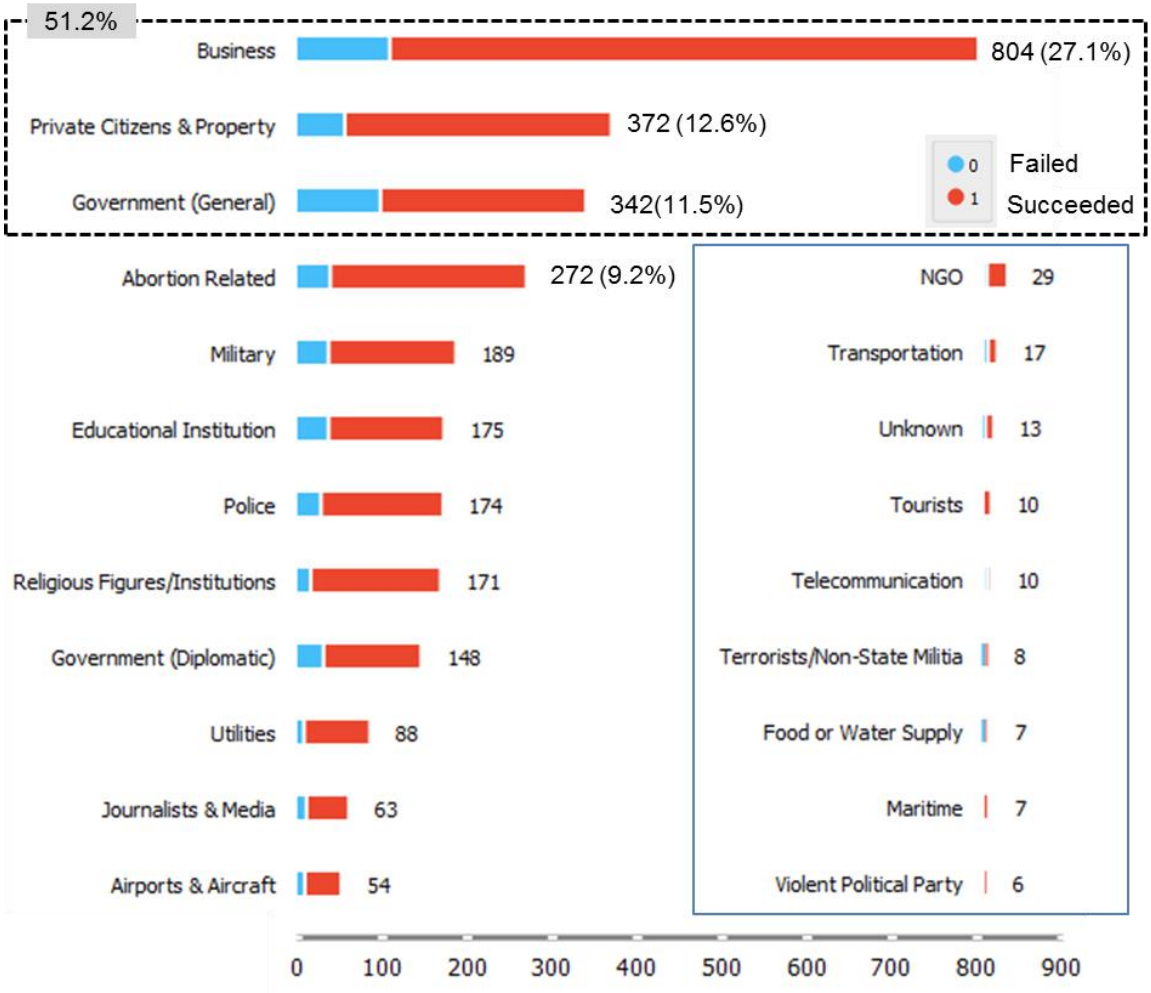

Figure 2: Distribution of U.S. attack targets (1970 – 2018).

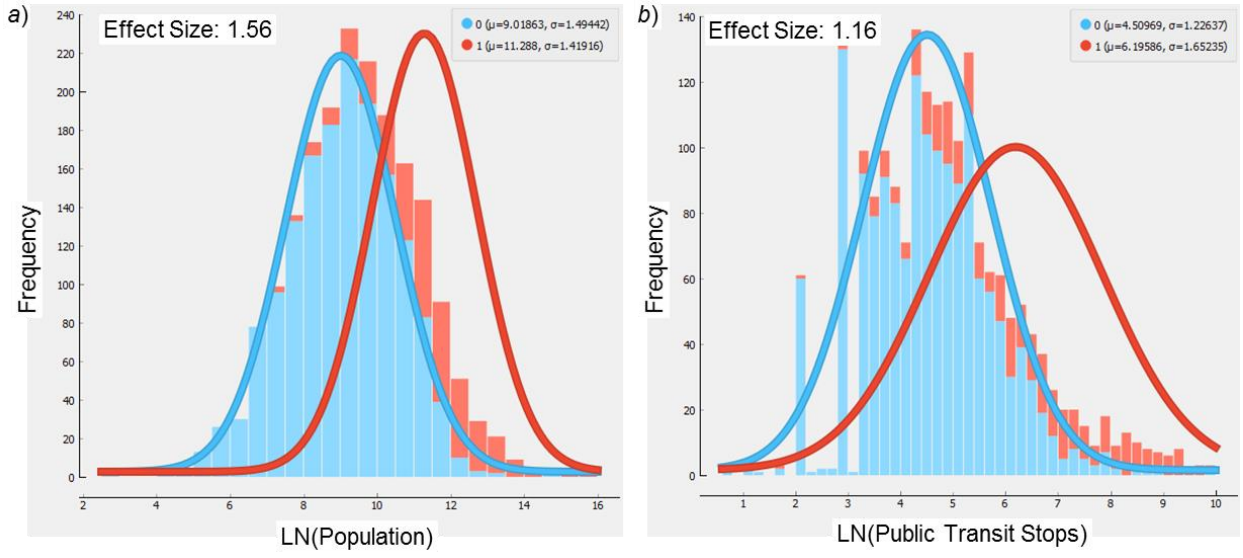

Figure 3: Effect size of the top two attributes.

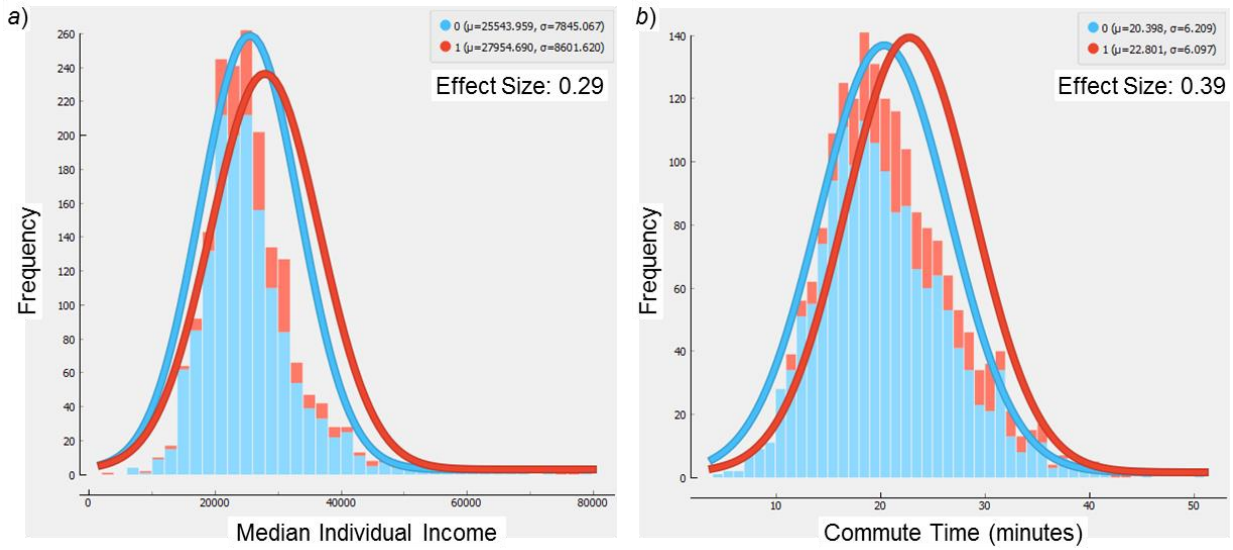

Figure 4: Effect size of two of the attributes that reduced the ML accuracy.

Table 1: Spatial Attribute Selection Based on EDA of the GTD.

Attribute	**Rationale**
Banks	Most frequently attacked target in the top category of "Business"
Fortune500	Several large companies were direct targets of previous attacks
Airports	A frequent target of attacks
Government	Government buildings are associated with attacks on governments
Military	Previous attacks on military or related facilities
Police	Previous attacks on law enforcement or related facilities
Courts	Previous attacks on courthouses or related judicial facilities
Worship	Previous attacks on places of worship
Colleges	Previous attacks on educational institutions
Hospitals	Previous attacks on healthcare facilities and related clinics
EPA	Attacks incited by Environmental Protection Agency (EPA) actions
Parks	Previous attacks related to the environment or animal rights

Table 2: Socio-Economic Attribute Selection from U.S. City Demographic Data

Attribute	Hypothesis
Accessibility	Number of transit access points can reduce reach to targets and escape options.
VMT	Busy target locations may divert attention from suspicious activities.
Commute Time	Commute time is a proxy measure of the degree of urbanization.
Housing Units	Residential density is proportional to the potential harm.
Age Distribution	Age demographic may be associated with target selection.
Race Distribution	Race demographic may be associated with target selection.
Gender	Gender demographic may be associated with target selection.
Limited English	Language demographic may be associated with target selection.
Disabled	Disability demographic may be associated with target selection.
Marital Status	Family demographic may be associated with target selection.
Family Size	Family demographic may be associated with target selection.
Veteran	Veteran demographic may be associated with target selection.
Education	Education demographic may be associated with target selection.
Income	Income demographic may be associated with target selection.
Employed	Employment demographic may be associated with target selection.
Home Ownership	Home ownership levels may be associated with target selection.
Home Value	Home values may be associated with target selection.
Rent	Rent burden may be associated with target selection.
Health Insured	Level of health insured may be associated with target selection.
Incorporated	A municipality with elected officials can lead to local grievances.
Township	A rural settlement with a local governance can lead to local grievances.
Average Elevation	Terrain or proximity to shores may be a spatial characteristic.

Table 3: Datasets and Sources

Dataset	Description	Records	Fields	Source
FTA	Facilities and vehicles by transit agencies	2,779	31	(FTA 2020)
APTA-F	Facilities maintained by transit agencies	501	43	(APTA-V 2020)
APTA-V	Vehicles maintained by transit agencies	6,936	36	(APTA-I 2018)
BTS	Population (2010) of U.S. cities and towns	38,186	19	(BTS 2019)
TIGER®	Shapefiles of U.S. county geographic	3,233	20	(USCB 2019)
GTD	Global Terrorism Database (1970 to 2018)	191,464	135	(START 2020)
GTD1993	Global Terrorism Database (1993)	748	135	(START 2020)
People	Demographics of U.S. cities and towns	108,796	81	(SimpleMaps 2020)
Worship	Places of worship in U.S. counties	49,328	39	(ASARB 2020)
Parks	Protected Areas Database	179,503	40	(USGS 2018)
Banks	Insured financial institutions and branches	87,847	30	(FDIC 2021)
Colleges	College and university campuses	5,865	27	(HIFLD 2017)
Courts	Courthouse locations in U.S. counties	3,719	26	(HIFLD 2017)
Government	Government buildings in U.S. counties.	1,378	24	(HIFLD 2017)
Police	Law enforcement locations in U.S. counties	23,486	36	(HIFLD 2017)
Fortune500	Headquarters of Fortune500 companies	500	22	(HIFLD 2017)
DoD	DoD facility locations in U.S. counties	664	9	(HIFLD 2017)
EPA	EPA facility locations in U.S. counties	77	14	(HIFLD 2017)
Hospitals	Hospital locations in U.S. counties	7,596	13	(HIFLD 2017)
Airports	Airport locations in U.S. counties	940	11	(HIFLD 2017)
NREL-T	Energy profile of U.S. counties (VMT)	3,142	44	(Ma, et al. 2019)
NREL-HU	Energy profile of U.S. counties (Housing)	3,142	44	(Ma, et al. 2019)

Table 4: Spatial attributes considered for the ML classification process.

Attribute	Description	Missing
Land_p_Worship	Density of places of worship in the county	152 (7%)
Land_p_College	Density of colleges in the county	487 (25%)
Land_p_Bank	Density of banks in the county	10 (0%)
Hospitals	Number of hospitals in the county	143 (7%)
VMT_p_Land	Vehicle-miles-traveled per square-mile of county land	53 (2%)
Police	Number of law-enforcement sites in the county	8 (0%)
HU_p_Land	Housing unit density (per square-mile)	0 (0%)
Courts	Density of courts in the county	18 (0%)
Parks_SQMI	Number of parks per square-mile in the county	230 (12%)
Land_p_Govt	Density of government buildings in the county	1,703 (89%)
F500	Number of Fortune 500 headquarters in the county	1,632 (85%)
Land_p_DoD	Density of DoD facilities in the county	1338 (70%)
Land_p_EPA	Density of EPA facilities in the county	1862 (97%)
Airports	Number of hospitals in the county	1,140 (59%)

Table 5: Predictive Performance Scores for Models Evaluated.

Model	AUC	CA	F1	Precision	Recall
MLP	0.893	0.881	0.873	0.872	0.881
XGB	0.889	0.877	0.590	0.752	0.492
GB	0.884	0.872	0.597	0.703	0.528
LR	0.882	0.878	0.866	0.869	0.878
RF	0.880	0.878	0.865	0.869	0.878
SGD	0.877	0.872	0.863	0.862	0.872
kNN	0.870	0.864	0.846	0.853	0.864
DT	0.842	0.869	0.860	0.859	0.869
NB	0.839	0.789	0.805	0.836	0.789
ADB	0.709	0.818	0.821	0.824	0.818
SVM	0.677	0.578	0.625	0.792	0.578
No-skill	0.497	0.818	0.736	0.669	0.818

Table 6: Attributes that contributed towards distinguishing the target class.

Attribute	$\chi 2$	d	Attacked
POP_LN	336.662	1.56	H
Stops_LN	193.519	1.16	H
VMT_p_Land_LN	179.867	0.93	H
Land_p_Bnk_LN	173.644	0.94	L
Hosp_LN	218.496	0.85	H
Land_p_Worship_LN	151.900	0.85	L
Limited_English_LN	156.542	0.83	H
Police_LN	154.506	0.76	H
Land_p_Coll_LN	90.005	0.75	L
Education_College+	108.602	0.66	H

7 Appendix

Table 7: Merging of Public Transit Datasets

			Missing Data for Cities		
Dataset	Operation	Cities	Stations	Stops	Vehicles
FTA	Extract and aggregate stations, vehicles.	2,291	1,789	N/A	395
APTA-F	Extract and aggregate stations, stops.	349	49	17	N/A
APTA-V	Extract and aggregate vehicles.	294	N/A	N/A	0
Merge-T	Combine FTA and APTA datasets by city.	2,291	1,743	2,006	393
Impute 1	Average stops/vehicle = 8.5	2,291	380	393	393
Impute 2	Average stops/station = 137.2	2,291	380	380	393
Transit	Drop records with no transit data defined	1,911	0	0	13

Table 8: Categories of Data Cleaning City Names

Category	**Examples**
Abbreviated	“Ft.” vs. “Fort”, “Mt.” vs. “Mount”, “St.” vs. “Saint”
Case Mix	Mixed case spellings (“Lagrange” vs. “LaGrange”, “La Place” vs. “LaPlace”)
Spelling	Incorrect spellings (“Tamarac” vs. “Taramac”, “Bellefontaine” vs. “Bellefonatine”
Extensions	Extra words such as “City of” or “Island of”, “Zuni Pueblo” vs. “Zuni”
Encoding	Names with non-English characters (“Española” vs. “Espanola”)
Alternate	Alternate names (“New York City” vs. “Manhattan”)

Table 9: Chronicle of Merging the Transit and Populated Places Datasets

			Missing Values	
Data	Operation	Cities	POP2010	Elevation
BTS	Extract city, county, state, FIPS, elevation, population.	38,186	11,125	0
De-Dupe	Merge records with duplicate city and state names	37,956	10,983	0
I-Search	Population of transit cities missing in the BTS	37	0	0
Merge	Fill missing population data.	37,956	10,946	0
Transit	Correct city names and merge with cleaned BTS	1,911	0	0

Table 10: Chronicle of Merging the GTD Dataset

Operation	**Description**	**Cities**
GTD	Extract city and state for all “United States” records	2,926
GTD1993	Extract city and state for all “United States” records	37
Stack	Append the GTD1993 records to the GTD records	2,963
Merge I	Merge the state abbreviation to form a unique key	2,963
Defined	Drop records for states that are “unknown”	2,957
Clean	Correct spelling of 26 city names	2,957
Aggregate	Group records by number of attacks in unique cities	819
Key	Create a unique key of city and state name	819
Merge II	Merge the county name and FIPS from BTS	819
Merge III	Merge with the combined transit and population data	1,911

Table 11: Attributes Selected from the “People” Dataset.

Attribute	Description	Missing
CITY	U.S. city name	0
COUNTY	U.S. county name	0
ST	State abbreviation	0
FIPS5	FIPS code (5 digits) uniquely identifying each county	0
Incorporated	“Y” if the place is incorporated, “N” otherwise	9 (0%)
Township	“Y” if the place is a township, “N” otherwise	9 (0%)
Military	“Y” if the place has a military related facility, “N” otherwise	9 (0%)
Commute Time	Average commute time in minutes	41 (2%)
Age Distribution	9 categories (Under 10, 10-19, 20’s, 30’s, etc.)	34 (1%)
Race Distribution	8 categories (Hispanic, Asian, Black, Multiple, Native, etc.)	80 (4%)
Gender	Male and Female proportion	80 (4%)
Limited English	Proportion of population considered to speak limited English	34 (1%)
Disabled	Percentage of population classified as “disabled”	34 (1%)
Marital Status	4 categories (Divorce, Married, Never Married, Widowed) %	80 (4%)
Family Size	Average number of family members in a household	34 (1%)
Veteran	Proportion of population that are veterans	34 (1%)
Education	7 categories (high school, bachelors, graduate, some, etc.) %	80 (4%)
Income	16 categories (poverty, dual, household, 10K, 100K+, etc.) %	36 (1%)
Employed	2 categories (labor force participation and unemployment rate)	36 (1%)
Home Ownership	Proportion of population owning a home	34 (1%)
Home Value	Average value of a home	77 (4%)
Rent	2 categories (Median Rent, Proportion of household income)	47 (2%)
Health Insured	Proportion of population with health insurance	34 (1%)