Journal of Public Transportation 24 (2022) 100009

Contents lists available at ScienceDirect

Journal of Public Transportation

journal homepage: www.journals.elsevier.com/journal-of-public-transportation

Using artificial intelligence to derive a public transit risk index

Raj Bridgelall

North Dakota State University, USA

ARTICLE INFO

Keywords:
Data Mining
Policymaking
Risk Assessment
Supervised machine learning
Terrorism tactics
Vulnerability assessment

ABSTRACT

A terrorist attack on the public transportation system of a city can cripple its economy. Uninformed investments in countermeasures may result in a waste of resources if the risk is negligible. However, risks are difficult to quantify in an objective manner because of uncertainties, speculations, and subjective assumptions. This study contributes a probabilistic model, validated by ten different machine learning methods applied to the fusion of six heterogeneous datasets, to objectively quantify risks at different jurisdictional scales. The risk index is purposefully simple to quickly inform a proportional prioritization of resources to make fair investment decisions that stakeholders can easily understand, and to guide policy formulation. The main finding is that the risk indices among public transit jurisdictions in the United States distribute normally. This result enables agencies to evaluate the quality of their risk index calculations by detecting an outlier or a large deviation from the expected value.

1. Introduction

The U.S. public transit system is operated by approximately 6800 transit agencies in tribal, rural, small urban, and large urban communities (APTA, 2020a). In 2018, the system handled almost 10 billion trips, representing a 21% increase from 1997. This increase in ridership was more than the 19% increase in the U.S. population during the same period. Recent studies indicate that millennials prefer public transportation over personal modes of mobility. This trend signals a continuation of ridership growth (TCRP, 2018).

Public transit systems worldwide have the common characteristics that they are large, open, populated, and highly interconnected. These properties make them both attractive and vulnerable to terror attacks (Needle and Renee, 1997). The consequences of a physical attack can lead to fatalities, injuries, and damage that can cripple the system's ability move people. After the September 11, 2001, attacks the U.S. government established the Transportation Security Administration (TSA) to increase airport security by controlling access to commercial aviation systems. A study shortly after found that there is a greater willingness to pay for risk reduction in air travel (Carlsson et al., 2004). However, a focus on aviation increases the risk that terrorists will divert their attention toward public transit systems. Yet, agencies pay little attention to physical security countermeasures because of uncertainties about threats and vulnerabilities. Unlike aviation, it is not practical to screen all public transit users (Jenkins and Butterworth, 2010). Instead, transit agencies tend to focus risk assessments on concrete safety issues such as susceptibility to natural hazards (FTA, 2019).

Terrorists are attracted to public transit systems because of the potential for catastrophic impacts from loss of capacity, easy access, the ability to act covertly within crowds, the ease of spreading agents and weapons throughout the system, and the convenience of access to quick escape routes (Bye et al., 2020). Attention generally turns to physical security in reaction to a recent event (Sunstein, 2003). By then, the task of conducting a risk assessment becomes daunting because it is impractical to consider all possible threats on a vast system. The most popular method of probabilistic risk assessment was borrowed from engineering to measure risk in direct proportion to levels of threat, vulnerability, and consequences (Stewart and Robert, 1997). However, such models are linear, and do not translate well to risk assessments that must consider the non-linearities of attack motives and behaviors. Subsequently, agencies must turn to empirical and subjective measures (Brown and Louis Anthony Cox 2011).

The lack of objectivity in empirical assessments lead to inconsistencies in risk level comparison, hindering effective investment prioritization, fair resource allocation, and policymaking. Hence, an objective quantification of risk requires an analytical approach to vulnerability assessment. However, it is not practical to conduct a vulnerability assessment by speculating on all possible tactics and attack targets. There is no model that can map attack tactics to likelihood of success in a public transit environment.

The goal of this research is to create a risk index that objectively quantifies the risk of attack on public transit systems. The Risk Analysis and Management for Critical Asset Protection (RAMCAP™) framework recommended by the U.S. Department of Homeland Security recommends (Brashear and William Jones, 2010) is a standard approach to risk assessment. The RAMCAP framework defines risk (R) as the product of threat (T), vulnerability (V), and consequences (C), commonly known as the TVC risk assessment. *Threats* are the specific modes

https://doi.org/10.1016/j.jpubtr.2022.100009

Available online xxxx

of attack that terrorists can use against their target. *Vulnerability* is the likelihood that an attack on the target will be *successful* when using the specific threat mode. *Consequences* depend on the target. The U.S. Federal Transit Administration (FTA) defines *consequences* as any harm or damage involving "injury, illness, or death; damage to or loss of the facilities, equipment, rolling stock, or infrastructure of a public transportation system; or damage to the environment" (FTA, 2019).

The objective of this research is to define a purposely simple risk index for public transit systems, derived using the RAMCAP framework. The approach uses machine learning (ML) models, a subset of artificial intelligence, to validate *threats* based on specific attack modes that public transit systems can be vulnerable to.

The contribution of this work is an objective, easy-to-calculate (simple), and scalable risk index that agencies can use to quantitatively compare the relative risks of attacks on public transit systems, at different levels of jurisdiction, for example at the local, county, state, or country levels. Ten complementary ML methods applied to the fusion of six datasets identifies RAMCAP *threat* and *vulnerability* factors associated with successful attacks. The main dataset is the Global Terrorism Database™ (GTD™), maintained by the National Consortium for the Study of Terrorism and Responses to Terrorism (START) at the University of Maryland (START, 2020). The other five datasets merged are from the U.S. Census Bureau, the Bureau of Transportation Statistics (BTS), the Federal Transit Administration (FTA), and two from the American Public Transportation Association (APTA).

The organization of the remainder of this paper is as follows: Section 2 surveys the literature to describe and compare related research that used ML methods to assess the security risks of public transit systems. Section 3 describes the data fusion, data preparation, and ML methods used to validate the variables in the proposed risk model. Section 4 describes the ranking of features and their contribution towards predicting successful attacks, and a statistical characterization of the risk index for cities where data is available. Section 5 discusses the results, implications, limitations, and generalizations of the analytical approach. Section 6 concludes the study by summarizing the approach and main findings, and hints briefly at extensions of this work for future contributions.

2. Literature review

Early work on risk perception by the public found that a complex mix of psychological, cultural, social, and institutional processes can amplify or attenuate the direct physical consequence of an event (Renn, et al. 1992). Both administrators and the public would similarly prioritize risk management (Carlsson, Daruvala and Jaldell, 2012). The public values the prevention of deaths from terrorism as much as they value the prevention of deaths from traffic accidents (Viscusi, 2009). Although not related to public transit security, Basuchoudhary and James (2018) used ML regression methods to predict the frequency and severity of terror attacks on countries (Basuchoudhary and James, 2018). They used the number of terror attacks from the GTD as a dependent variable and sourced the independent variables from other cross-national databases that contained measures of political, socioeconomic, cultural, ethnic, and religious divisions within a country. A key finding is that random forest (RF) provided the best predictions on the test data. The RF method builds a diverse set of decision trees for final voting by incorporating randomness in both the tree-splitting variable subset and the data sampling.

Regain and David (2017) applied *association rule learning* to the GTD and found that in suicidal terrorist events there are strong associations between the country of both the attacker and victims (Regian and David, 2017). Although not strictly ML, Grant and Stewart (2015) applied a probability model to the GTD data and found that attacks using improvised explosive devices (IEDs) in the United States are less likely to succeed than those perpetrated in other parts of the Western world (Grant and Stewart 2015). Similarly, Kirisci (2018) conducted a probabilistic assessment of the GTD data and noticed that states with strong bureaucracies tend to be better protected from domestic terrorism (Kirisci, 2018).

There were a few studies about public transit security, but they did not involve the use of ML models. Loukaitou-Sideris et al. (2006) determined that the open and accessible nature of public transit systems has attracted more frequent attacks (Loukaitou-Sideris et al., 2006). A recent study found that passengers and operators experienced assaults in 85% and 75% of the agencies surveyed, respectively (Bye et al., 2020). Yet, only 25% of transit agencies have implemented a security-risk-reduction program that they considered to be effective. Fiondella et al. (2012) found that a mass-transit passenger screening checkpoint will drastically reduce the flow of passengers (Fiondella, et al. 2012). From a study of bus operations in the Los Angeles, California system, Pearlstein and Wachs (1982) found that crime increase was directly proportional to the growth in transit ridership, in both space and time (Pearlstein and Wachs, 1982).

As transit agencies struggle to balance security countermeasures with the openness and attractiveness of their systems, they find that a partnership with intelligence and law-enforcement agencies becomes inevitable and crucial. One study found that transit security professionals most often cite the deployment of a uniformed patrol as the most effective method of deterring terrorist attacks (Needle and Renee, 1997). Another study suggests that deploying cameras and algorithms to recognize human behaviors can be an effective countermeasure for transit systems (Joshua et al., 2009). Some agencies have developed smartphone apps that let passengers record and report any suspicious activity (Sneider, 2016). However, not all transit agencies can afford to deploy such tools across the entire system. Therefore, it is important to be able to quantify relative risks, such as using a risk index, to focus resources and attention where they are most needed.

General risk assessment frameworks, such as the RAMCAP can guide the development of a risk index but they are subjective. Game theory was a popular approach used to model the interactions among adversarial agents to inform resource allocation for airport security (Jiang et al., 2014). Delle et al. (2014) developed a general Bayesian Stackelberg game model to inform the dynamic allocation of security resources in uncertain domains and applied it to a case study of Metro trains in Los Angeles, California (Delle et al., 2014). Based on the literature search, the work presented in this paper will be the first to use ML models to guide the development of a risk index for public transit security. This is not surprising because there is extraordinarily little overlap among the fields of physical security, public transportation, and machine learning.

3. Method

The analytical workflow involves three layers of procedures as shown in shown in Fig. 1. The data fusion layer prepares and cleans the data for the ML layer. The processing layer uses a variety of ML methods to systematically identify RAMCAP *threat* and *vulnerability* features that are associated with successful attacks. The next subsections describe the procedures in each layer of the framework.

3.1. Data fusion

Table 1 describes the general structure of the six datasets used to build the machine learning models. The GTD™ contains attributes and text narratives that describe each terror attack (START, 2020). Attributes include the attack date, location, perpetrator group, tactics, weapons, consequences, data source, variable descriptions, and text narratives. The attack outcome variable "success" has a value of 0 when attacks failed.

The Topologically Integrated Geographic Encoding/Line (TIGER/Line™) database from the U.S. Census Bureau encodes map data in a

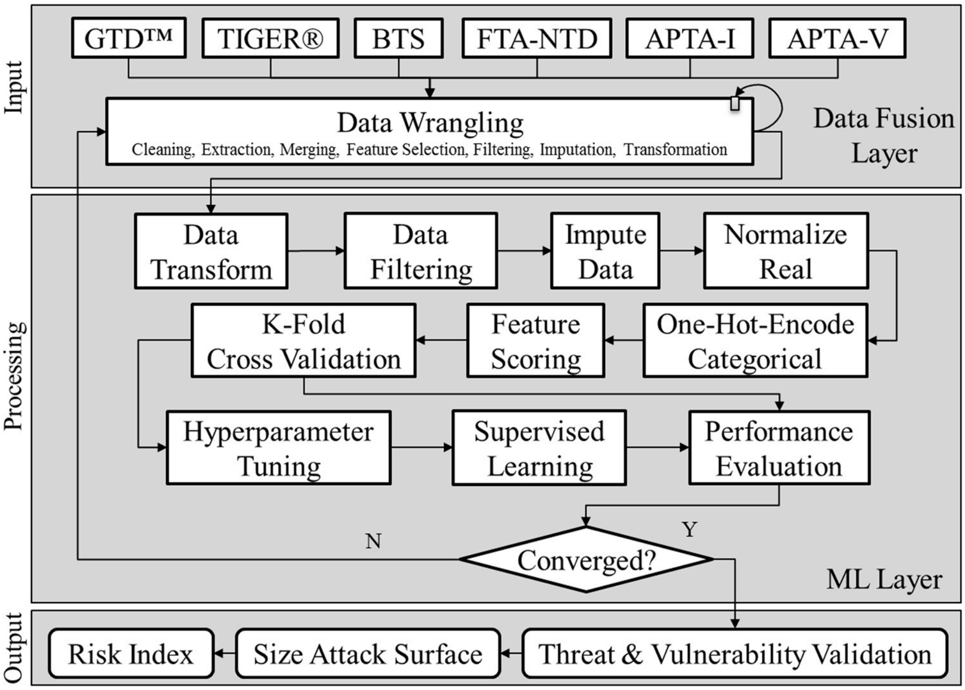

Fig. 1. The analytical workflow of the procedures used.

standard "shapefile" format (USCB, 2019). The database subset for U.S. counties encodes the boundaries as polygons. The associated data tables contain information about each county, such as their names, state, Federal Information Processing Systems (FIPS) codes, centroid geospatial coordinates, land area, and water area. The "Populated Places" database from the BTS contains 2010 census population and elevation data for many populated places in the United States, including county, state, geospatial coordinates, and FIPS codes (BTS, 2019).

The National Transit Database (NTD) from the FTA has many data subsets that contain information about transit agencies, assets, fares, expenses, ridership, and services (Anon, 2020). The extracted "Facilities and Stations" dataset from the 2018 Annual Data Tables lists the amount, type, and age distribution of passenger stations that each agency used to provide bus, rail, and ferry services. The APTA database contains transit statistics in several data subsets, including those that do not report to the FTA-NTD (APTA, 2020b). The two APTA datasets are the 2019 Vehicle Database (APTA-V) and the 2016 Infrastructure Database (APTA-I). They augmented or reconciled data from the FTA dataset.

3.2. *Data wrangling*

The series of methods used to prepare the fused dataset for machine learning was data extraction, data cleaning, data merging, feature selection, data filtering, data imputation, and variable transformation. *Data extraction* identified and loaded only portions of large heterogeneous datasets that were relevant to the analysis. *Data cleaning* identified and replaced erroneous or missing values for all attributes. *Data merging* combined datasets by using either unique attribute keys or geospatial intersections to fill missing values and to add attributes that improve the classification accuracy. This process discovered errors when records failed to merge, thus provided a feedback loop, as shown in Fig. 1, to repeat earlier procedures until achieving convergence. *Feature selection* identified attributes in the datasets that were relevant, sufficiently dispersed, and without too many missing values. *Data filtering* removed outliers to reduce model generalization and accuracy. *Data imputation* populated missing data so that other features of the observation could add information to the ML process without bias. *Variable transformation* reduced the distribution skew of variables,

Table 1
Datasets used in the ML process.

Database	Description	Version and Structure
GTD™	Open-source database hosted by the University of Maryland. Contains information on terrorist incidents around the globe since 1970. Variables include the attack location, attack type, target type, weapons used, causalities, consequences, and many text narratives (START, 2020).	2019 Release: 135 fields; 191,474 records. Separate 1993 records (35).
TIGER	Provided by the US Census Bureau. A geographic information systems (GIS) map file (shapefile) and data tables that include state and county names, FIPS codes, land area, and water area (USCB, 2019).	2019 Release: 11 fields; 3108 counties
BTS	Provided by the BTS. Contains the 2010 census population and average county elevation relative to sea level (BTS, 2019).	May 2019 Release: 19 fields; 38,186 records
FTA	Provided by the FTA. The "Facilities and Stations" database provide a distribution of the number and types of passenger facilities by the decade that they were completed. Includes the number of vehicles operated by agencies to provide service at full capacity (Anon, 2020).	2018 Release: 21 fields; 4964 records.
APTA-I	Provided by APTA. Contains a list of all significant infrastructure from 617 transit agencies in the United States and Canada. Includes the number of types of passenger facilities used to provide bus, rail, and ferry services (APTA, 2020b).	Infrastructure 2016 Release: 43 fields; 501 records.
APTA-V	Provided by APTA. Contains information about fleet size and type from 160 transit agencies (APTA, 2020b).	Vehicle 2019 Release: 36 fields; 67,94 records.

Table 2
Features selected or extracted for ML process.

Variable	Description	Type	Dispersion	Missing %
City	City of attack	Text	–	0
State	State of attack	Text	–	0
County	County of attack	Text	–	0
FIPS5	Combined state and county FIPS code	Text	–	0
A_LAND	Land area (square miles)	Integer	1.700	0
A_WATER	Water area (square miles)	Integer	1.540	0
ELEV	Average elevation of county (feet)	Integer	1.970	0
POP_2010	2010 Census of county population	Integer	1.274	0
Lat	Latitude geospatial coordinates	Real	0.125	0
Lon	Longitude geospatial coordinates	Real	0.206	0
DY	Incident day	Cat	0.599	0
MO	Incident month	Cat	0.541	0
WTC	Weapon type category	Cat:Table 3	1.240	0
ATC	Attack type category	Cat:Table 3	1.310	0
TTC	Target type category	Cat:Table 3	2.350	0
WTO	Weapon type ordinated by frequency of use	Ordinal	0.140	0
ATO	Attack type ordinated by frequency of use	Ordinal	0.170	0
TTO	Target type ordinated by frequency	Ordinal	0.210	0
Fac_Decade	Number of transit facilities in attack decade	Integer	1.571	34
Fac_2018	Number of transit facilities in 2018	Integer	0.770	53
Vehicles	Number of transit vehicles	Integer	1.151	31
success	Attack success (1) or failure (0)	Binary	0.452	0

Table 3
Categorical codes in the GTD for attack mode features.

WTC	ATC	TTC
1. Biological	1. Kill	1. Business
2. Chemical	2. Shoot	2. Government (general)
3. Radiological	3. Explode (bomb)	3. Police
4. Nuclear	4. Hijack	4. Military
5. Firearms	5. Imprison (hostage)	5. Abortion (related)
6. Explosives	6. Kidnap (hostage)	6. Aviation
7. Fake (weapons)	7. Vandalize (facility)	7. Diplomats
8. Fire (incendiary)	8. Assault (unarmed)	8. Educational Institutions
9. Melee (fight)	9. *Unknown*	9. Sustenance (food/ water)
10. Vehicle		10. Media (News outlets)
11. Sabotage (equipment)		11. Maritime
12. *Other*		12. Non-Gov. Organization
13. *Unknown*		13. *Other*
		14. Person (individuals)
		15. Worship (related)
		16. Telecommunication
		17. Militias
		18. Tourists
		19. Transportation (non-aviation)
		20. *Unknown*
		21. Utilities
		22. Extremists (political parties)

normalized continuous variables, and encoded categorical variables into binary representations that were suitable for the ML models.

Table 2 and Table 3 summarizes the final feature set of the cleaned and merged tables. The amount of dispersion for each variable indicates the relative amount of their variability or information content. Table 3 lists the categorical codes for the weapon type category (WTC), attack type category (ATC), and target type category (TTC).

3.3. *Machine learning*

This subsection provides a succinct review of the ML methods used. The next subsections describe the scoring methods used to rank feature importance, the ML methods used to assess RAMCAP factors that contribute to the prediction of successful attacks, a cross-validation method to improve model generalization, and measures of model performance.

3.3.1. *Data processing*

The data transformation procedure reduced the skew of continuous variables to improve the performance of some models by applying a shifted log transformation (Géron 2019). With each decision loop in the model development framework, the data filtering procedure removed attributes and outlier observations that did not contribute to the predictive performance of the ML models. The data imputation procedure used one of several known methods to fill missing values for features that contributed to the ML performance (Géron 2019). To improve the performance of some ML methods, the feature normalization procedure converted the values of continuous variables to the [0,1] range. Some models cannot work directly with categorical variables. Therefore, the *one-hot-encoding* procedure created one new attribute per value in the category such that the new attribute has a value of 1 if the attribute is present and 0 otherwise (Potdar et al., 2017).

3.3.2. *Feature scoring*

Datasets with many features, relative to the number of instances, can decrease the performance of ML methods because features may be noisy, irrelevant, or contain redundant information (Yu and Liu, 2003). There are a variety of methods available to score features based on the amount of information they contribute towards class separation. Table 4 provides a short description of each method and a reference that details their theory of operations and implementation. Each method tends to compensate for some weakness of the other, so the rankings can differ (Wang et al., 2010).

3.3.3. *K-fold cross validation*

The technique of k-fold cross validation supports a measure of the *generalized* performance of the tuned ML models by using the entire dataset. Rather than setting aside a portion of the data as a test set, k-fold cross validation partitions the data into *k* approximately equal size subsets, each to be used only once for validation and the union of the remainder for training the model. The average of the selected performance metric across all *k* validation cycles is the measure of generalized performance. Géron (2019) provides a hands-on treatment of the method (Géron 2019).

3.3.4. *Supervised learning*

There are many different types of supervised ML models and each tend to fit a type of dataset better than others (James, et al. 2013). This section provides a brief overview of 10 different types of ML models

Table 4
Feature scoring methods.

Method	Description	References
Information Gain	The expected amount of entropy reduction. A decrease in entropy (uncertainty) based on the presence of another variable will increase information.	(Yu and Liu, 2003)
Gini Decrease	A measure of the inequality among values of a frequency distribution based on their statistical dispersion. A value of zero and one represents perfect equality and inequality, respectively, of a variable and the class distributions.	(Han et al., 2016)
ANOVA	Analysis of Variance (ANOVA) measures the difference between average values of the feature in different classes by using the F distribution.	(Agresti, 2018)
Chi-Squared	Measures a dependency or association between the feature and the class by using a chi-square statistic.	(Wang et al., 2010)
FCBF	Fast Correlation Based Filter (FCBF) measures entropy and accounts for redundancy among features without doing pairwise correlations.	(Yu and Liu, 2003)

used to predict attack outcome. Table 5 groups the models into four broader categories: Tree-based Methods, Statistical Models, Decision Boundaries, and Learned Functions. Many of the models have hyperparameters that require user adjustment to maximize performance. Table 5 includes a brief description of each model, the hyperparameters (HP) that need adjustment, overall advantages (A) and disadvantages (D), and a reference that provides more detail about their theory of operations and implementation.

The no-skill classifier provides a baseline measure of performance for a classifier that simply predicts the dominant class each time. The procedure will also evaluate the performance of a stacked metaclassifier, which combines the outputs of several base classifiers. Géron (2019) provides detailed descriptions of all the models and their theory of operation (Géron 2019).

3.3.5. *Performance evaluation*

Various performance metrics guide the hyperparameter tuning to yield the best *generalized* performance. The five metrics used are classification accuracy (CA), precision (Pc), recall (Rc), F1-score (F1), and the area under the curve (AUC) of the receiver operating characteristic (ROC). The definition of each metric uses the true positive (TP) and false positive (FP) rates of the predictions. CA is the proportion of correct predictions, whether positive or negative. Pc is the proportion of correct positive predictions where Pc = TP/(TP + FP). Rc is the

Table 5
Overview of ML models used.

Category	Model	Algorithm & Hyperparameters	Advantages and Disadvantages
Tree-Based Methods	Decision Tree (DT)	Tree node splitting. HP: Minimum number of instances in leaves (N), and minimum size of subsets (S) (Aggarwal, 2015).	A: Simple to interpret and to visualize. D: Tends to overfit, resulting in low predictive power on new data.
	Random Forest (RF)	Build full trees for forest voting from a bootstrapped dataset with randomly selected attributes. HP: Number of trees (*N*) and minimum size of subsets (S) (Breiman, 2001).	A: Combines the simplicity of decision trees with less tendency of overfit, thereby improving prediction accuracy. D: incomplete trees diminish insights.
	AdaBoost (AB)	Sequentially build improved shallow trees for forest voting. HP: Number of estimators (*N*), learning rate (*R*), boosting algorithm, and regression loss function (James, et al. 2013).	A: selects only those features that improve predictive power, hence, reducing the computational burden for datasets with very large dimensionality. Less sensitive to overfitting. D: Sensitive to the presence of outliers and data with high incoherence.
	Extreme Gradient Boost (XGB)	Sequentially build improved models that fit the errors of previous models. HP: Number of estimators (N), learning rate (*R*), maximum tree depth (*S*), loss function (Chen and Guestrin, 2016).	A: efficient and good performance on large datasets; inherently supports missing values. D: sensitive to hyperparameter selection; requires manual intervention to achieve the best configuration for a given dataset.
Statistical Models	*k*-Nearest Neighbors (*k*-NN)	Determine the class of an instance based on the majority class of its k nearest neighbors. HP: Number of neighbors (*k*), Distance method (Géron 2019).	A: simplicity of method. D: sensitive to a skewed class distribution. The computational intensity grows exponentially with the number of instances and attributes.
	Naïve Bayes (NB)	HP: none (Aggarwal, 2015).	A: fast and simple method. D: poor performance when attributes are not independent.
	NoSkill	A trivial model that predicts the dominant class each time. Used only as a baseline to compare the performance score of skilled classifiers (Géron 2019).	N/A
Decision Boundaries	Logistic Regression (LR)	Establish a decision boundary by using a logistic function to maximally separate classes. HP: Regularization function and strength (C), and probability threshold (Aggarwal, 2015).	A: inherits many of the advantages of linear regression; precisions are easy to make. D: sensitive to noise in the data such as outliers and incorrectly classified instances. Model fitting may fail to converge if there are many highly correlated features.
	Support Vector Machine (SVM)	Establish a decision boundary by finding a multidimensional hyperplane to maximally separate classes. HP: Kernel type, cost (C), and regression loss (ε) (Platt, 2000).	A: high accuracy with low computational complexity. D: sensitive to noisy data and multidimensional planes that lack clear boundaries.
Learned Functions	Stochastic Gradient Descent (SGD)	An optimization technique that fits a linear multivariate function to the data. It works best when all features are scaled. HP: loss function, learning rate method and parameters (Aggarwal, 2015).	A: an efficient technique on large datasets. D: sensitive to feature scaling; many hyperparameters; and the true minima may not be achieved because the gradient is only an approximation.
	Artificial Neural Network (ANN)	A weighted multilayer linear network that represents a function. HP: Hidden layer neurons (N), Solver type, regularization parameter (α), number of iterations (I) (Aggarwal, 2015).	A: accuracy improves with use and feedback about classification accuracy. D: requires many training examples to improve classification accuracy.

proportion of all the true positives predicted where Rc = TP/(TP + FN). F1 is the harmonic mean of Pc and Rc where F1 = TP/(TP + α) and α = (FN + FP)/2. The ROC is a plot of the TP rate against the FP rate of a binary classifier, as a function of its discrimination threshold. The AUC is the area under the ROC curve where 1.0 represents perfect classification performance. As described by Krawczyk (2016), the AUC is more complex to calculate but it is best suited for class imbalanced data (Krawczyk, 2016). Fawcett (2006) provides further insights into the interpretation of each performance metric (Fawcett, 2006).

3.4. Risk index

The ML models validated the *threat* factor of the RAMCAP framework by ranking features of historical attacks that predicted successful attacks. The ML models also rank key features of the public transit system that increase the *vulnerability* factor of the RAMCAP framework. Those key features are the number of transit vehicles (vehicles), number of transit facilities (Fac_2018), number of facilities in the attack decade (Fac_Decade), and the population of the county (POP_2010) that each transit system served.

A risk index for public transit systems and infrastructure would be proportional to the size of the *attack surface*. The terminology “attack surface” in security refers to the number of vulnerable locations in a system and does not necessarily refer to a contiguous “surface” in space. This model defines the size of the attack surface *S* for the public transit system as the sum of passenger access points—the number of passenger transit facilities and fleet vehicles.

The proposed risk index (*R*) is a simple computation that is proportional to the size of the attack surface (*S*) and to the number of historical attacks (*A*) in the jurisdiction of the transit agency such that

$$R=ln(S \times A) \tag{1}$$

where ln is the natural log. The factor *S* encapsulates the *threat* and *vulnerability* factors of the RAMCAP model that the ML methods validate. It is important to note that if none of the ML models find that the identified threat factors (attack modes) and vulnerability factors (attack surface size) contribute towards the predictability of successful attacks, then S = 0. Similarly, S = 0 if there were no attacks in a jurisdiction. That is, the factor *A* characterizes the probability of an attack in the jurisdiction, based on historical data. The frequency of attacks in a jurisdiction may be associated with some attractiveness of that location, even though not causally related to the potential for affecting public transit systems there.

4. Results

This section presents the results of the feature ranking, machine learning, and quantification of the risk index based on the models described previously.

4.1. Feature scoring

Table 6 shows the top 30 features, sorted in the order ranked by the ANOVA scoring method. The table shows the relative rankings of all five methods. It is evident that all methods produce consistently high ranking for the most lethal attack type categories of “Kill” and “Explode” but rank geospatial coordinates such as “Lat” and “Lon” near the bottom. Table 7 shows the pairwise correlation of the ranking by all methods. It is evident the Gini, Information Gain, ANOVA, and χ^2 methods are all highly correlated, and this increases the confidence of their relative strength of association with successful attacks. The FCBF method had the least correlation among methods but had the highest correlation with ANOVA.

The top ranked categorical values, with 1 being the most important, indicated that the RAMCAP *threat* factors of killing or vandalizing government interests with the use of explosives, guns, or fire are most strongly associated with successful attacks. Indeed, the defined *attack surface S* of public transit systems would be vulnerable to such attacks. Fig. 2 shows that adding features beyond the top ranked 30 scored by the ANOVA method did not appreciably improve the prediction performance of the model.

Factors of the defined attack surface size, which were the number of facilities in the decade of an attack, the number of public transit vehicles, and the number of facilities in 2018 ranked 9, 13, and 14, respectively. Those features, which are components of the RAMCAP

Table 6
Top 30 features ordered by the ANOVA ranking.

Variable	Info. Gain	Gini	ANOVA	χ2	FCBF
ATC = Kill	4	2	1	1	1
ATC = Explode	3	4	2	5	10
TTC = Government	7	5	3	2	3
ATC = Vandalize	5	6	4	4	11
ATC = Shoot	6	7	5	3	2
WTC = Biological	9	9	6	6	13
WTO	1	1	7	7	9
WTC = Guns	11	12	8	9	14
Fac_Dec_Log	10	10	9	10	4
TTC = Business	14	15	10	12	18
Land_Log	12	11	11	14	21
ATO	2	3	12	8	12
Veh_Log	15	14	13	11	25
Fac_2018_Log	13	13	14	21	20
WTC = Melee	18	20	15	13	15
WTC = Fake	28	23	16	15	22
ATC = Assault	26	24	17	16	23
ATC = Imprison	22	25	18	17	17
TTC = Worship	24	26	19	19	24
TTC = Utility	25	27	20	20	19
Elev_Log	21	21	21	25	32
TTC = Diplomat	30	28	22	22	27
POP_2010_Log	20	19	23	18	31
Lat	19	18	24	27	28
DAY	29	29	25	24	37
ATC = Kidnap	27	30	26	26	5
Lon	23	22	27	23	33
WTC = Sabotage	33	32	28	28	30
TTC = Aviation	37	33	29	29	36
ATC = Hijack	31	34	30	30	7

Table 7
Correlation of scoring methods.

Method A	Method B	Correlation
Gini	Info. Gain	0.982
ANOVA	χ2	0.962
ANOVA	Gini	0.907
ANOVA	Info. Gain	0.897
Gini	χ2	0.885
Info. Gain	χ2	0.873
ANOVA	FCBF	0.673
FCBF	χ2	0.664
FCBF	Info. Gain	0.645
FCBF	Gini	0.582

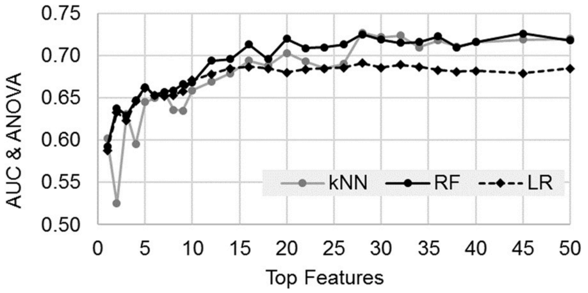

Fig. 2. AUC performance for top ANOVA ranked features.

Table 8
ML model performance and tuned hyperparameters.

Model	AUC	CA	F1	Pc	Rc	Hyperparameters
Stack	0.741	0.857	0.820	0.848	0.857	*Base*: ANN, kNN, RF, LR, AB. *Agg*: LR
kNN	0.732	0.851	0.809	0.837	0.851	k = 25 (odd), *Distance*: Euclidean
RF	0.706	0.844	0.809	0.815	0.844	$N = 25$, $S \geq 5$
XGB	0.702	0.844	0.913	0.855	0.979	*Loss*: LR, $S = 6$, $R = 0.1$
ANN	0.689	0.836	0.778	0.798	0.836	$N = 200$, *Activation*: ReLu, *Solver*: SGD, $\alpha = 10^{-4}$
AB	0.683	0.813	0.803	0.796	0.813	$N = 100$, $R = 0.1$, *Boost*: SAMME, *Loss*: Linear
LR	0.682	0.840	0.778	0.823	0.840	*Reg*: Ridge, $C = 1$
SGD	0.677	0.838	0.775	0.814	0.838	*Loss*: LR. *Reg*: L1 norm. *Learn*: IS (0.01, 0.6)
NB	0.656	0.727	0.744	0.765	0.727	None
Tree	0.582	0.810	0.790	0.779	0.810	$N = 10$, $S = 5$
SVM	0.561	0.745	0.745	0.746	0.745	*Kernel*: Sigmoid, $C = 1$, $\varepsilon = 0.1$
No-Skill	0.497	0.833	0.757	0.694	0.833	None

vulnerability factor, were within the model improvement region, suggesting that they are positively associated with attack success.

4.2. Machine learning

The ML models provide a measure of association based on predictive power. The next sections summarize the results of the ML performance evaluation and assurance of individual model optimization by hyperparameter tuning.

4.2.1. Performance evaluation

Table 8 indicates that the best models can predict attack success with an accuracy of 85% based on the *threat* and *vulnerability* features. The top-ranking classifier had an AUC score of 0.75, which is significantly greater than the AUC score of 0.50 for no-skill classifier. The stacked metaclassifier performed best as expected, but only slightly better than the kNN and RF methods. In general, the results indicate that models that used loss functions or probabilities perform worst on this fused dataset than methods that incorporate majority voting. For instance, the kNN method and the RF method seeks a majority vote among k nearest neighbors and N randomly grown decision trees, respectively. The stacked metaclassifier seeks a majority vote among its base classifiers.

4.2.2. Hyperparameter tuning

Fig. 3 shows the result of hyperparameter tuning for the top performing individual classifiers, namely kNN and RF. The hyperparameter N represents the number of trees for the RF, and the number of nearest neighbors for kNN. The AUC and CA scores are the average measures from a 10-fold cross validation that used all features to train and test the models. To minimize bias, the cross-validation used stratified sampling to equally represent the minority target class in each fold. The trend indicates that the performance measures increase in an asymptotic manner as the parameter increases.

The fluctuations reflect the randomness from fold creation in the cross-validation process. The trend suggested that an empirical choice for the number of RF trees would be 25. Similarly, considering the asymptotic trend, an empirical choice for the number of nearest neighbors would be 25. A combinatorics grid-search method derived the best value for all the hyperparameters shown in Table 8.

4.3. Risk index assessment

Based on the ML outcome, an assessment that follows the RAMCAP framework indicates that an attack on public transit systems using the top-ranking attack modes will result in successful attacks with high likelihood. Simply put, the defined attack surface would be vulnerable to explosives and fire. Armed perpetrators can harm passengers or take hostages. The feature ranking and predictive performance of the ML models also indicated that the size of the attack surface is associated

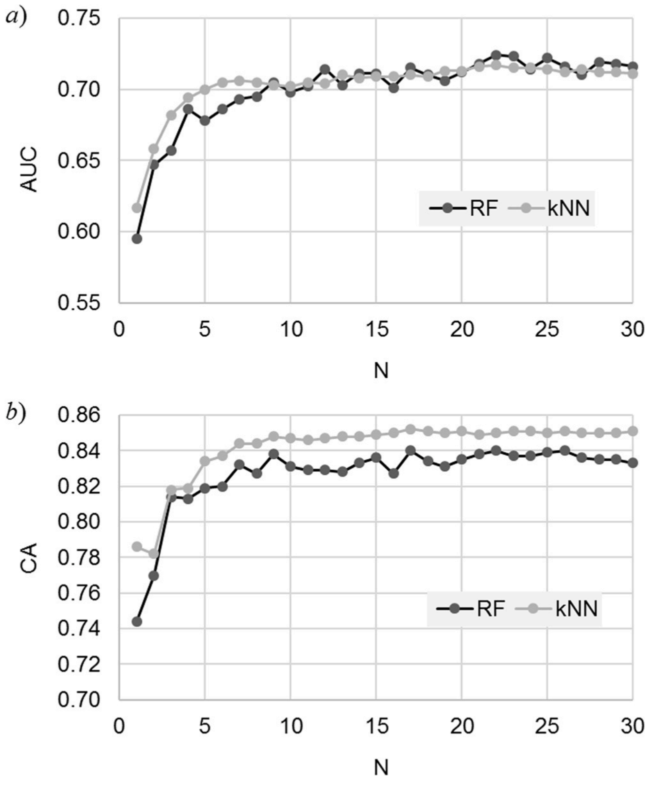

Fig. 3. Hyperparameter tuning for the RF and kNN models based on a) AUC and b) CA.

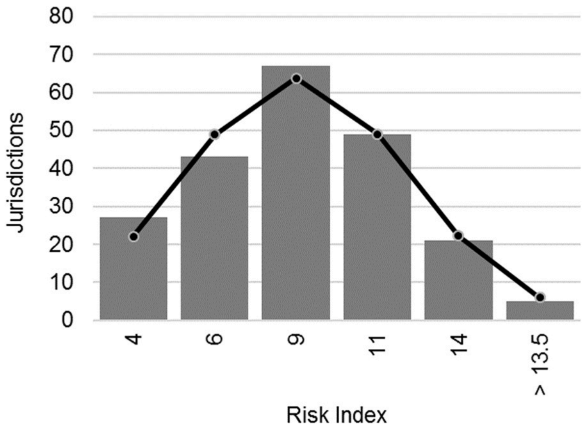

Fig. 4. Distribution of risk indices.

with successful attacks. Given this validation by the ML models, the next subsections quantify the risk index and their distribution for all GTD cities where information on public transit systems is also available in the FTA and APTA databases. There were 212 jurisdictions in the GTD that met this criterion.

4.3.1. Normal distribution

The bar chart of Fig. 4 shows a histogram of the calculated risk indices and the line plot shows the best fit Gaussian function.

The best fit is determined by solving the optimization problem:

$$\underset{X_i}{\text{minimize}} \quad e = \sum_{i=1}^{B} (H_{\text{i}} - D_{\text{i}})^2$$

subject to $\alpha > 0, \sigma > 0,$ and $N \geq B \geq 4$

$$\text{where } D_{\text{i}} = \frac{\alpha}{\sqrt{2\pi\sigma^2}} e^{-\frac{(X_i-\mu)^2}{2\sigma^2}}, \text{ i} = 1, 2, \ldots, \text{B} \tag{2}$$

The counts for values within interval X_i of bin i are H_i. The values evaluated at interval X_i is D_i where the amplitude α, mean μ, and variance σ^2 minimize the sum-of-squares (SOS) error e.

The goodness-of-fit of the solution is determined by a Pearson's chi-squared test where the statistic is

$$\chi_k^2 = \sum_{i=1}^{B} \frac{(H_{\text{i}} - D_{\text{i}})^2}{D_{\text{i}}} \tag{3}$$

The degrees-of-freedom (*df*) associated with the chi-squared statistic is *k*, which is the number of histogram bins minus the three parameters *α*, μ, and *σ*. The *B* histogram bins must be at least four so that the *df* can be at least unity, and the upper bound cannot exceed the number of samples *N*. The probability *p* that the chi-squared statistic obtained is at least as large as the expected value is the area under the chi-squared distribution curve

$$p = \int_{\chi_k^2}^{\infty} \frac{1}{2^{k/2}\Gamma(k/2)} x^{k/2-1} e^{-x/2} dx \tag{4}$$

where $\Gamma(k/2)$ is the gamma function in mathematics. The computed probability, which is also known as the p-value, indicates that the null hypothesis, which is that the distribution follows the fitted Gaussian, cannot rejected when the value is greater than 0.05 (Agresti, 2018). The p-value associated with this test was 0.823, which indicates that the statistical test could not reject the hypothesis that the distribution follows a Gaussian.

4.3.2. Biased spatial dispersion

Fig. 5 shows how the risk indices distribute spatially across the continental United States. Attacks appear to be broadly dispersed, albeit with some overall bias towards major cities along the west coast, northeast, and southern United States.

The map pattern indicates that low- and high-risk indices tend to associate with cities that have low and high populations, respectively, but not exclusively. This outcome of the objective ML process agrees with the subjective intuition. Table 9 summarizes the cities with public transit risk indices ranking in the top 10.

5. Discussion

The scalability of the risk index stems from its suitability for use at any spatial *scale* to quantify security risks from microscopic to macroscopic levels. For example, agencies can compare risk indices among townships, cities, counties, or states. Townships within a state that did not experience terrorist attacks will have undefined indices. However, aggregating up to the county level can produce a defined risk index by aggregating attacks across townships within a county. The Gaussian distribution of risk indices suggests that there is a structure that encapsulates some natural relationship among the *attack surfaces* (*S*) and *attack likelihood* (*A*) among cities. That is, the central tendency of the risk indices indicates a similarity relationship among agencies. The distribution also identifies outliers. Therefore, agencies can use the distribution as a quality check to gauge the likelihood of their calculated risk index, based on a Gaussian distribution.

The attack frequency and attack surface are not independent because there is a small positive correlation of 0.5 between them. This is an indication that the attractiveness of a target is inherent in the size of the transit system. There is also a weak association between the size of the attack surface and the success of an attack. From Table 6, the variables associated with public transit size ranked between 9 and 14 in their association with successful attacks. The weak correlation between the size of the attack surface, the attack frequency, and their association with successful attacks will result in a weak non-linear amplification of the risk index. Hence, the risk index inherently incorporates the destruction of property as a potential consequence. There is also a positive and stronger correlation of 0.67 between population size and attack surface. Hence, the risk index also inherently incorporates harm to people as a potential consequence.

The risk index quantifies the risk of an attack, regardless of "hardening" from security measures currently in place. Agencies could add a "risk reduction factor" to account for "hardening" measures relative to others. However, doing so would require much more data, which would

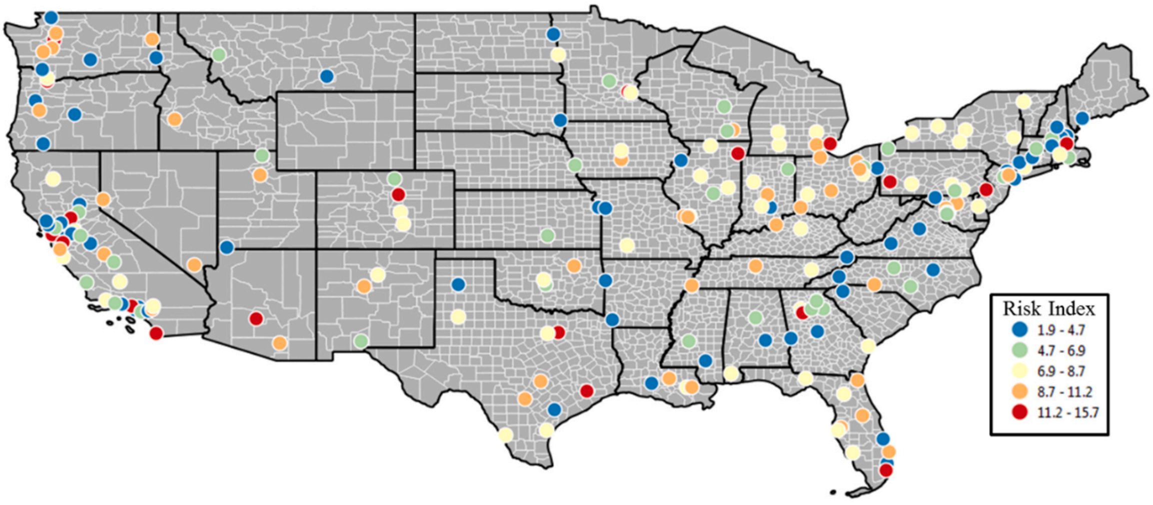

Fig. 5. Spatial distribution of risk indices.

Table 9
Cities with public transit risk indices ranking in the top 10.

Jurisdiction	State	Pop_2010	Attacks	Success	Vehicles	Facilities	Surfaces	Risk Index
Brooklyn	NY	2565,635	469	367	13,189	428	13,617	15.7
Los Angeles	CA	3792,621	112	96	4508	18,568	23,076	14.8
Washington	DC	601,723	85	64	3930	11,366	15,296	14.1
Chicago	IL	2695,598	63	50	4792	11,533	16,325	13.8
Miami	FL	399,457	87	79	1396	9013	10,409	13.7
San Francisco	CA	805,235	98	79	1517	4242	5759	13.2
Seattle	WA	608,660	41	33	4790	8406	13,196	13.2
Denver	CO	600,158	24	23	1703	10,368	12,071	12.6
Houston	TX	2099,451	21	18	2659	10,103	12,762	12.5
Boston	MA	617,594	18	16	3252	8201	11,453	12.2

diminish the benefit of a simple first-pass risk calculation that could preclude more expensive follow-up assessments. Also, adding a risk reduction factor could lead to misjudgments and possible complacency as terrorists continuously adapt their tactics and methods to exploit vulnerabilities that managers might miss. Adversaries may also review the literature on current risk management strategies and change their tactics accordingly.

In addition to providing data-driven justification for RAMCAP *threat* and *vulnerability* factors, the ML processes produced some insights about characteristics of the dataset. Only four of the models had an AUC performance above 0.70 with a corresponding classification accuracy of 85%, suggesting that all strategies were similarly affected by high noise in the dataset. Noise can manifest as spillage into the margins that may otherwise naturally separate clusters. Furthermore, noise can contaminate the homogeneity of clusters. The lack of clean boundaries is a likely explanation for the poor performance of SVM, which seeks a clear separating hyperplane in the *global* structure of feature space as explained previously. This finding is similar to those of Wang et al. (2010) where SVM proved to be the worst performer on a highly imbalanced dataset (Wang et al., 2010). Conversely, the majority vote type algorithms of kNN and RF performed best because of their reliance on *local* similarity and randomized global searches, respectively, as explained previously.

Risk managers can generalize the framework to derive risk indices for other types of facilities. For example, risk indices can be developed for railroad, pipeline, and bridge networks based on the size and accessibility of their attack surface, and the vulnerability of their attack surfaces to explosives or fire, which are the top ranking variables associated with successful attacks. Users of the model need not repeat the predictive ML modeling because this work already assessed the relevance of the *threat* and *vulnerability* factors based on the RAMCAP framework. That is, to quantify risk, users need only to define and determine the size of their attack surface and the likelihood of attacks based on the historical attack frequency. This simple model enables low-cost first-order risk analysis that can use existing public domain data.

This work does not suggest that the simple model could replace a more complex and extensive analysis that requires internal data, expert knowledge, and situational awareness across all agency products. One limitation of this framework is that it does not define a risk index for facilities within jurisdictions where terrorist attacks did not occur. The reason for this limitation is that the ML models cannot predict the attack frequency for a location with no attack history because of a lack of relevant attributes. Therefore, it is important to note that an undefined risk does not translate to zero risk, nor does it suggest that a jurisdiction has no features that could attract terrorist activities.

6. Conclusions

Public transit systems are vast, open, populated, and critically important to the vitality of nations. Hence, they are attractive targets of terrorism. Although it is impossible to predict the timing of terrorist attacks, trends from the Global Terrorism Database (GTD™) show that they are ongoing and widespread. Hence, response preparation through regular risk assessments are critical. Yet, most agencies defer investments and lack policies on countermeasures to physical attacks. The uncertainty of risks and the subjectivity of risk assessments can discourage a focus on physical security. Furthermore, it is impractical and cost prohibitive to implement security checkpoints in public transit systems. Therefore, transportation managers and decision makers can benefit from a more objective, focused, and probabilistic approach to guide fair prioritization, investment decisions, and policymaking.

This research developed a simple transit security risk index that is objective, easy to calculate, and scalable to multiple levels of jurisdictions. It may be tempting to underestimate the power of the index due to its simplicity. However, it is important to consider that the risk index is a probabilistic function of attack likelihood and the size of the *vulnerable* attack surface, both of which the machine learning (ML) models validated as relevant factors in predicting successful attacks. The size of the attack surface in this case is the number of fleet vehicles, stations, platforms, and other facilities that expose passengers to harm.

The framework applied ML to a fusion of the GTD and databases from the U.S. Census Bureau, the Bureau of Transportation Statistics, the Federal Transit Administration, and the American Public Transportation Association. Ten complementary ML models provided different levels of predictive performance. The best models were *k*-nearest neighbors and random forest. They could predict attack success with an accuracy of 85%.

The public transit risk indices distribute normally for jurisdictions in the GTD where information was available to compute their attack surface and attack frequency. This finding suggests that the proposed risk index encapsulates a structural relationship between risk, attack likelihood, and the size of public transit systems. Attack risk was spatially dispersed, albeit a bit biased towards the largest U.S. cities with huge public transit systems. This objective finding from the statistical framework matches expectations. The Gaussian distribution provides a means to assess the quality of the index by detecting outlier calculations or a large deviation from the expected value.

Future work will explore the use of unsupervised methods of machine learning to produce additional insights about underlying structural relationships among terrorist incidents. Another study is underway to apply ML to features of attacked versus spared locations to understand factors that may have contributed to their attractiveness.

Disciplines

Defense and Security Studies | Emergency and Disaster Management | Peace and Conflict Studies | Terrorism Studies | Transportation.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Aggarwal, Charu C., 2015. Data Mining. Springer International Publishing, New York, New York.

Agresti, Alan, 2018. Statistical Methods for the Social Sciences. Pearson, Boston, Massachusetts.

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

APTA, 2020a. 2020 Public Transportation Fact Book. 71st.. American Public Transportation Association, Washington, D.C.. 〈https://www.apta.com/wp-content/uploads/APTA-2020-Fact-Book.pdf〉.

APTA, 2020b. Transit Statistics. American Public Transportation Association June 24. Accessed 24 June, 2020. 〈https://www.apta.com/research-technical-resources/transit-statistics/〉.

Basuchoudhary, Atin, James, T.Bang, 2018. Predicting terrorism with machine learning: lessons from "predicting terrorism: a machine learning approach. Peace Econ. Peace Sci. Public Policy 24 (4). https://doi.org/10.1515/peps-2018-0040

Brashear, Jerry P., William Jones, J., 2010. Risk Analysis and Management for Critical Asset Protection (RAMCAP Plus). Wiley Handbook of Science and Technology for Homeland Security. John Wiley & Sons Inc, pp. 1–15. https://doi.org/10.1002/9780470087923.hhs003

Breiman, Leo, 2001. Random forests. Mach. Learn. 45 (1), 5–32. https://doi.org/10.1023/A:1010933404324

Brown, Gerald G., Jr. Louis Anthony Cox, 2011. How Probabilistic Risk Assessment can Mislead Terrorism Rrisk Analysts. Risk Analysis: An International Journal 31 (2), 196–204. https://doi.org/10.1111/j.1539-6924.2010.01492.x

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

Bye, Patricia, Ernest, R., Frazier, Sr, 2020. Transit Security Preparedness. TCRP Synthesis 146. Transit Cooperative Research Program (TCRP), Transportation Research Board (TRB), Washington, D.C.. https://doi.org/10.17226/25764

Joshua, Candamo, Shreve, Matthew, Goldgof, Dmitry B., Sapper, Deborah B., Kasturi, Rangachar, 2009. Understanding transit scenes: a survey on human behavior-recognition algorithms. IEEE Trans. Intell. Transp. Syst. (Inst. Elect. Elect. Eng. (IEEE)) 11 (1), 206–224. https://doi.org/10.1109/TITS.2009.2030963

Carlsson, Fredrik, Daruvala, Dinky, Jaldell, Henrik, 2012. Do administrators have the same priorities for risk reductions as the general public. J. Risk Uncertain. 45 (1), 79–95. https://doi.org/10.1007/S11166-012-9147-3

Carlsson, Fredrik, Olof, Johansson-Stenman, Peter, Martinsson, 2004. Is transport safety more valuable in the air. J. Risk Uncertain. 28 (2), 147–163. https://doi.org/10.1023/B:RISK.0000016141.88127.7C

Chen, Tianqi, and Carlos Guestrin. 2016. XGBoost: A Scalable Tree Boosting System. The 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 785–794. doi:10.1145/2939672.2939785.

Delle, Fave, Francesco, Maria, Albert Xin, Jiang, Zhengyu, Yin, Chao, Zhang, Milind, Tambe, Sarit, Kraus, Sullivan, John P., 2014. Game-theoretic patrolling with dynamic execution uncertainty and a case study on a real transit system. J. Artif. Intell. Res. 50, 321–367. https://doi.org/10.1613/jair.4317

Fawcett, Tom, 2006. An introduction to ROC analysis. Pattern Recognition Letters Elsevier, pp. 861–874. 〈http://people.inf.elte.hu/kiss/11dwhdm/roc.pdf〉.

Fiondella, Lance, Swapna, S.Gokhale, Lownes, Nicholas, Accorsi, Michael, 2012. Security and Performance Analysis of a Passenger Screening Checkpoint for Mass-Transit Systems. 2012 IEEE Conference on Technologies for Homeland Security ((HST)). Institute of Electrical and Electronic Engineers (IEEE),, Waltham, Massassachutes, USA, pp. 312–318. https://doi.org/10.1109/THS.2012.6459867

FTA, 2019. Sample Safety Risk Assessment Matrices for Bus Transit Agencies. U.S. Department of Transportation. Federal Transit Administration (FTA), Washington, D.C., pp. 21.

Géron, Aurélien, 2019. Hands-on Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems. O'Reilly Media, Sebastopol, California.

Grant, Matthew, Mark G., Stewart, 2015. Probabilistic Risk Assessment for Improvised Explosive Device Attacks that Cause Significant Building Damage. Journal of Performance of Constructed Facilities 29 (5), B4014009. https://doi.org/10.1061/(ASCE)CF.1943-5509.0000694

Han, Hong, Guo, Xiaoling, Yu, Hua, 2016. "Variable Selection Using Mean Decrease Accuracy and Mean Decrease Gini Based on Random Forest.". The 7th IEEE International Conference on Software Engineering and Service Science ((ICSESS)). Institute of Electrical and Electronic Engineers (IEEE), pp. 219–224. https://doi.org/10.1109/ICSESS.2016.7883053

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

Jenkins, Brian M., Bruce, R.Butterworth, 2010. Explosives and Incendiaries Used in Terrorist Attacks on Public Surface Transportation: A Preliminary Empirical Analysis. MTI Report WP 09-02. Mineta Transportation Institute Publications, San Jose, California MTI Report WP 09-02. 〈https://scholarworks.sjsu.edu/mti_publications/134/〉.

Jiang, Albert Xin, Jain, Manish, Tambe, Milind, 2014. Computational game theory for security and sustainability. J. Inform. Process. 22 (2), 176–185. https://doi.org/10.2197/IPSJJIP.22.176

Kirisci, Mustafa, 2018. Fighting Terrorism Through an Effective Bureaucracy. Beyond the Horizon ISSG, Brussels. 〈https://behorizon.org/fighting-terrorism-through-an-effective-bureaucracy/〉.

Krawczyk, Bartosz, 2016. Learning from imbalanced data: open challenges and future directions. Prog. Artif. Intell. 5 (4), 221–232. https://doi.org/10.1007/s13748-016-0094-0

Loukaitou-Sideris, Anastasia, Brian, D.Taylor, Camille, N.Y.Fink, 2006. Rail transit security in an international context: lessons from four cities. Urban Aff. Rev. 41 (6), 727–748. https://doi.org/10.1177/1078087406287581

Needle, Jerome A., Renee, M.Cobb, 1997. Improving Transit Security. Transit Cooperative Research Program (TCRP), Transportation Research Board (TRB), Washington, D.C.

Pearlstein, Adele, Wachs, Martin, 1982. Crime in public transit systems: an environmental design perspective. Transportation 11 (3), 277–297. https://doi.org/10.1007/BF00172653

Platt, John C., 2000. Probabilities for SV Machines. In: Alexander, J.Smola, Peter, J.Bartlett, Dale, Schuurmans, Bernhard, Schölkopf (Eds.), in Advances in Large Margin Classifiers. MIT Press, Cambridge, Massachusetts, pp. 61–74.

Potdar, Kedar, Taher, S.Pardawala, Pai, Chinmay D., 2017. A comparative study of categorical variable encoding techniques for neural network classifiers. Int. J. Comput.Appl. 175 (4), 7–9. https://doi.org/10.5120/ijca2017915495

Regian, J.Wesley, and David A.Noever. 2017. Generative Representation of Synthetic Threat Actors for Simulation and Training. Interservice/Industry Training, Simulation, and Education Conference (I/ITSEC). 10.

Renn, Ortwin, William, J.Burns, Kasperson, Jeanne X., Kasperson, Roger E., Slovic, Paul, 1992. The social amplification of risk: theoretical foundations and empirical applications. J. Soc. Issues 48 (4), 137–160. https://doi.org/10.1111/J.1540-4560.1992.TB01949.X

Sneider, Julie. 2016. Focus on Transit Security. Progressive Railroading, January.

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〉.

Stewart, Mark, Robert, E.Melchers, 1997. Probabilistic Risk Assessment of Engineering Systems. Springer, Netherlands.

Sunstein, Cass R., 2003. Terrorism and Probability Neglect. J. Risk Uncertain. 26 (2), 121–136. https://doi.org/10.1023/A:1024111006336

TCRP, 2018. Understanding Changes in Demographics, Preferences, and Markets for Public Transportation. Research Report 201, Transportation Research Board (TRB). Transit Cooperative Research Program (TCRP), Washington, D.C.. https://doi.org/10.17226/25160

USCB, 2019. TIGER/Line Shapefiles Technical Documentation. United States Census Bureau (USCB), Washington, D.C., pp. 138.

Viscusi, W.Kip, 2009. Valuing Risks of Death from Terrorism and Natural Disasters. J. Risk Uncertain. 38 (3), 191–213. https://doi.org/10.2139/SSRN.1359221

Wang, Huanjing, Taghi, M.Khoshgoftaar, Gao, Kehan, 2010. A Comparative Study of Filter-Based Feature Ranking Techniques. 2010 IEEE International Conference on Information Reuse & Integration. Institue of Electrical and Electronic Engineers (IEEE)., Las Vegas, Nevada. https://doi.org/10.1109/IRI.2010.5558966

Yu, Lei, and Huan Liu. 2003. Feature Selection for High-Dimensional Data: A Fast Correlation-Based Filter Solution. The Twentieth International Conference on Machine Learning (ICML-2003). Washington, D.C. 856–863. https://www.aaai.org/Papers/ICML/2003/ICML03–111.pdf.