social sciences

MDPI

Article

An Application of Natural Language Processing to Classify What Terrorists Say They Want

Raj Bridgelall

Department of Transportation, Logistics, and Finance, College of Business, North Dakota State University, Fargo, ND 58108, USA; raj.bridgelall@ndsu.edu

Abstract: Knowing what perpetrators want can inform strategies to achieve safe, secure, and sustainable societies. To help advance the body of knowledge in counterterrorism, this research applied natural language processing and machine learning techniques to a comprehensive database of terrorism events. A specially designed empirical topic modeling technique provided a machine-aided human decision process to glean six categories of perpetrator aims from the motive text narrative. Subsequently, six different machine learning models validated the aim categories based on the accuracy of their association with a different narrative field, the event summary. The ROC-AUC scores of the classification ranged from 86% to 93%. The Extreme Gradient Boosting model provided the best predictive performance. The intelligence community can use the identified aim categories to help understand the incentive structure of terrorist groups and customize strategies for dealing with them.

Keywords: counterterrorism; machine learning; risk modeling; sustainable societies; text mining

check for updates

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

Academic Editor: Nigel Parton

Received: 3 November 2021
Accepted: 10 January 2022
Published: 13 January 2022

1. Introduction

Knowing the aims of terrorist attacks can help the intelligence community devise different strategies to deal with the perpetrators. Hence, the main goal of this research is to identify terrorist aims and classify them into a finite set of categories. This research defines aims of terrorists as distinctly different from the root causes of terrorism. Aims are the agenda or desired outcome of an attack (e.g., provoke a US attack upon Muslims, total war upon "non-believers", and the creation of a global Caliphate promoted by Al-Qaeda) whereas root causes explain the birth of terrorists (e.g., the desire to establish the Islamic rule worldwide) (Burke 2021).

There has been no internationally agreed-upon definition of terrorism (Cassese 2006). Reich (1990) submitted that no single perspective can explain the complex and diverse phenomenon of terrorism. The dominant paradigm of research in counterterrorism assumes that terrorists are rational actors that attack civilians for political ends (Géron 2017). Therefore, as rational actors, terrorists must have an agenda or aim. However, systemic models that attempt to explain aims involve a difficult tradeoff between generalizability and accuracy (Hastie et al. 2016).

Abrahms (2008) suggests that there is no other question than "What terrorists really want?" that is more fundamental for devising effective strategies in counterterrorism. Different aims may require different strategies to deal with the perpetrators. Hence, classifying attack aims instead of root causes brings another perspective to help advance the counterterrorism agenda. The recent work of Wong et al. (2021) strongly suggested that such knowledge is integral to Government's policy responses in countering the threat of terrorism and violent extremism. Consequently, the objectives of this research are to:

1. Apply natural language processing (NLP) techniques to extract features from the stated motive narrative of terrorist attacks to identify perpetrator aim categories (PACs).
2. Validate the effectiveness of the PAC classification by evaluating the predictive performance of 11 different machine learning (ML) models applied to a different narrative

field, the event summary. The significance of evaluating multiple types of ML models is that no single model type can best represent all datasets, which is an accepted tenant in the practice of ML (Aggarwal 2015).

Aside from identifying root causes, there has been no counterterrorism research to classify the general aims of perpetrators. Hence, this research contributes an artificial intelligence (AI) framework to help us understand what terrorists want based on a "revealed preference" gleaned from the narratives of what terrorists say they want. The framework combined NLP and ML techniques, both of which are subfields of AI (Aggarwal 2015), to aid human cognition in deriving the PACs. Intelligence communities can use the framework to classify motive narratives and guide decisions about the types of resources and mitigation efforts needed to best address the PAC identified for an event. Examples of resources are experts in certain fields of psychology, sociology, policymaking, law-enforcement, churches, and education. Knowing the PAC can guide more cost-effective selection of the appropriate resources much earlier in the counterterrorism effort.

The remainder of this article is organized as follows: Section 2 reviews related works to characterize terrorist motives. Section 3 applies NLP to extract PACs and Section 4 validates the classification using ML techniques. Section 5 discusses the significance of the classification results and the effectiveness of the AI methods. Section 6 recaps the approach, the current findings, and points to future work to examine group activity patterns and their association with the identified PACs.

2. Literature Review

Krieger and Meierrieks (2011) cautioned that the phenomenon of terrorism is too complex to be reduced only a single root cause and panacea. Maszka (2018) found that although theoretical models in terrorism studies are useful, greater accuracy requires greater complexity. Recent analysis found that academic efforts to study the effectiveness of counterterrorism strategies are lacking (Balestrini 2021). Scholars in the literature on terrorism studies tend to analyze motives from one of three perspectives: (1) root causes, (2) planned impacts on the public, and (3) aims of the perpetrators. The next three subsections summarize the literature on each analytical perspective. The fourth subsection discusses related work that applied ML to better understand various aspects of terrorism.

2.1. Root Causes

Treistman (2021) observed that research on the causes of terrorism tends to focus on macroscopic trends at the national level without considering contextual factors such as social exclusion that affect individuals. An analysis of 44 terrorist organizations listed by the European Union found that 45% of them were politically motivated, nearly the same proportion was social-revolutionary motivated, while religiously motivated groups accounted for about 10% of the total (Rothenberger and Müller 2015). An earlier study highlighted racism as a root cause (Björgo 1993). Araújo et al. (2020) found that the perception of immigrants as a threat affected social attitudes towards foreign groups. A more recent study discussed the role of shame and revenge as a root cause in terrorism (Cottee 2021). Rigterink (2021) also recently found evidence that revenge explained the increase in terrorist attacks after the assassination of a terrorist leader. Van Um (2011) characterized root causes of terrorism based on need, such as the desire for unit-reinforcement and self-enrichment. Coccia (2018) presented statistical evidence that fatalities associated with terrorism are in regions where high population growth rates may result in inequality, subsistence stress and deprivation.

Awareness of the need for self-enrichment increased with the proliferation of modern communications and social media (Höflinger 2021). On the other hand, user-generated data from social media can help analysts to profile users but the noise generated and data size proliferation can pose significant challenges (Bilal et al. 2019). Höflinger (2021) posits that modern communications technologies has blurred the boundaries between the actions of terror organizations and individuals.

2.2. Planned Impacts

Monahan and Valeri (2018) posits that terrorists aim to maximize fear to achieve their strategic objectives. Terrorists seek attention and even claim responsibility for their acts (Morley and Leslie 2007). Masuku et al. (2021) found that terrorists use propaganda to expand their operations and seek sympathizers. News of terrorist attacks can threaten audiences' perceptions of societal safety (Tamborini et al. 2020). Hence, crowded urban areas are typically prime targets (Coaffee 2009). Canetti et al. (2021) found that hate-motivated attacks elicited stronger reactions from the public. One study found that among the tactics used, mass shootings and suicide bombings tend to result in the most causalities (Arce 2019).

2.3. Terrorist Aims

For studies in counterterrorism strategies, it is crucial to understand the decision processes of terrorists if we wish to plan countermeasures (Jaspersen and Montibeller 2020). Research found that deterrence requires knowledge about terrorist aims to inform proactive policies that can deter attacks (Enders and Su 2007). Kydd and Walter (2006) suggest that the aims of terrorists are to change minds through five strategies: attrition, intimidation, provocation, spoiling, and outbidding. Cottee and Hayward (2011) characterized terrorist motives from the perspective of three desires—the need for excitement, ultimate meaning, and glory. Kurtulus (2017) found evidence that terrorists aim to mobilize their constituency, avenge their fallen associates, and physically destroy their perceived enemies. Abrahms (2008) offered that the most common strategies to fight terrorism are to diminish the aims via strict no-concession policies or via appeasement. However, the aims of terrorist leaders and their operatives may be misaligned. In an analysis of terrorist propaganda videos, Abrahms et al. (2017) found that the operatives of terrorist leaders commit more indiscriminate violence than their leaders favor.

2.4. ML Applications

Luo and Qi (2021) used a random forest model to rank factors in terrorist attack risk and found that human-loss ranked highest. Huamaní et al. (2020) applied ML to the global terrorism database (GTD) to predict terrorist attacks worldwide. In related work, Guo et al. (2007) introduced a visualization environment to identify patterns in the GTD. Huff and Kertzer (2018) used ML to predict how the public classified incidents as terrorism based on language used in media coverage. In similar work, Das and Das (2019) used NLP to extract paraphrases from a large untagged text corpora to discover labels of crime reports. Mashechkin et al. (2019) proposed some language-independent algorithms to extract information from the Internet that contained terrorist or extremist patterns. Khalifa et al. (2019) applied association mining to the GTD to understand the nature of terrorist attacks in Egypt.

Burnap and Williams (2014) found that statistical modeling and ML was effective in classifying "hate speech" on Twitter as measured by an F1 score of 0.95. Bassetti et al. (2018) used a Generalized Mixed Effects Regression Tree analysis to analyze economic correlates of Islamist political violence. Uddin et al. (2020) compared the performance of ML models in predicting attack success, involvement of suicide, weapon type, attack type, and region. Canhoto (2021) leveraged ML for counterterrorism by detecting and preventing money laundering, which is a key tool in terrorist operations. Hao et al. (2019) applied random forest to the GTD for spatio-temporal pattern discovery of terrorism incidents on the Indochina Peninsula. Mishra et al. (2020) applied directed graphs to the GTD for network relationship discovery of terrorist activities. Feng et al. (2020) applied Extreme Gradient Boosting to the GTD for predict casualty prediction from terrorist attacks.

3. Methodology

This section introduces the dataset used, the development of an empirical topic classification method to extract relevant features from motive narratives, and the predictive ML models used to validate the association of event summaries with the theorized PACs.

3.1. Data

The 2019 release of the Global Terrorism Database (GTD) contained more than 201,183 records of terrorist events from 1970 to 2019 (START 2020). Of the 135 fields containing information about each event, 42 remained after removing fields that were missing more than 35% of the data. Eighteen of the 42 fields contained textual data, but only two were relevant to the NLP procedures. The two relevant text fields were "summary" which contained a brief description of the terrorist event, and "motive" which briefly described the motive, if known, stated in media reports. Therefore, the reported motive is akin to a "stated preference" of what the perpetrators say was the aim of their attack.

Figure 1 summarizes the attack frequency aggregated by five-year blocks from 1970 to 2019, and by the portion with motives known. Only 10,747 or 5.3% of the records described reported motives, based on responsibility claims by the perpetrators. The trend was that the number of records with known motives increased to a stable level after 1998. Therefore, this analysis focused on records with claimed motives for the 20 years spanning 1999 through 2019. This filter yielded 9460 records, which represented only 12% fewer records than all records containing known motives.

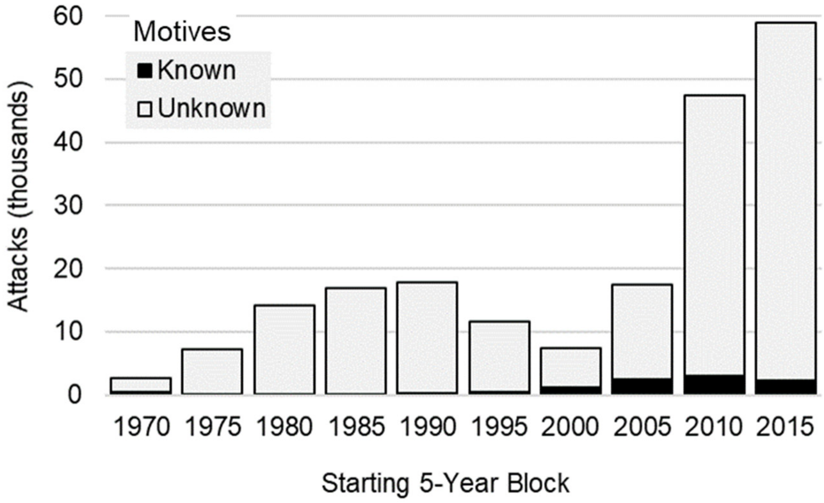

Figure 1. Attacks aggregated by 5-year blocks and motive information from 1970 to 2019.

3.2. Topic Classification

NLP is a subset of the field of Artificial Intelligence (Aggarwal 2015). Topic modeling is an NLP method that requires both art and science to identify and quantify the mix of topics within a document (Lane et al. 2019). Latent Dirichlet Allocation (LDA) is the most prominent among several techniques currently available (Padmaja et al. 2018). LDA identifies topics based on a multinomial distribution of words and assumes that each document covers a distinct topic. The available topic modeling algorithms do not accommodate user inputs about application context or topic names. After applying LDA to the motive narrative field of the GTD, the algorithm identified topics based on a group of frequent words. Empirical analysis of the word groups by the author revealed that they tended to associate with weapon types, attack types, and perpetrator groups, each with some mix of probabilities. Hence, LDA did not help the author to identify PACs.

Given the above deficiencies, the author invented an empirical topic classification (ETC) technique that combined NLP methods with human cognition to identify word features that resulted in a more meaningful identification of PACs. Figure 2 shows the workflow, including custom and existing NLP procedures. The data processing layer extracted motive narratives from the selected year range of records. The author then initialized a list of potential PACs based on expertise, data sources listed in the GTD, keywords gleaned from the literature, as well as the keywords selected from the motive narrative.

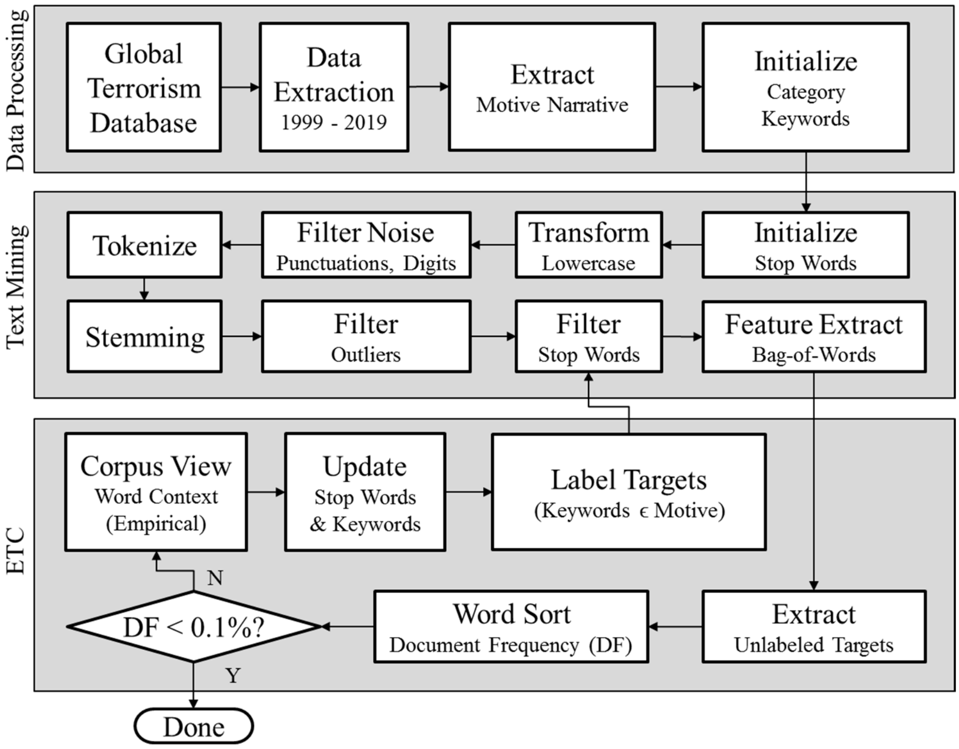
Data Processing
Global Terrorism Database
Data Extraction 1999 - 2019
Extract Motive Narrative
Initialize Category Keywords
Text Mining
Tokenize
Filter Noise Punctuations, Digits
Transform Lowercase
Initialize Stop Words
Stemming
Filter Outliers
Filter Stop Words
Feature Extract Bag-of-Words
ETC
Corpus View Word Context (Empirical)
Update Stop Words & Keywords
Label Targets (Keywords ϵ Motive)
N
DF < 0.1%?
Word Sort Document Frequency (DF)
Extract Unlabeled Targets
Y
Done

Figure 2. Workflow for empirical topic classification.

The text mining layer then applied NLP procedures to normalize the text and extract relevant word features. The lowercase transformation prevented word redundancy and the tokenize procedure extracted individual words into a feature vector that represented each narrative of a record. The stemming procedure reduced feature dimensionality by transforming all inflected forms of a word to their lexical stem (Jones and Willett 1997). The stop words filter eliminated frequently used words such as “the”, “on”, “at”, “which”, “and”, “but”, to reduce bias noise from a lack of meaning. The outlier word filter further reduced noise by allowing the downstream word sort procedure to highlight meaningful words that appear in many of the documents. The bag-of-words model extracted the document frequency (DF) of words based on the number of documents in the corpus that contained those words.

With each iteration of the procedural flow, the BOW procedure of the ETC framework helped human cognition to assign a PAC to the unlabeled targets based on DF rank, which rapidly reduced the number of unassigned document targets with each iteration. This strategy also helped human cognition to more easily identify additional stop words and keywords associated with a PAC. The Corpus View procedure sorted the word features of the unlabeled documents by their DF and highlighted the context of a selected word across all documents. This strategy helped human cognition to focus the ETC on the most frequent words, and to achieve an exponential convergence of label assignments. The iterations stopped when the most frequent remaining words appeared in less than 0.1% of the documents. The rational for the termination condition was that infrequent words added noise to the ML process, and that could lead to model over-fitting in the subsequent ML layer.

In summary, the ETC workflow was a machine-aided human decision process that leveraged the domain knowledge and experience of the author to help assign appropriate PACs to motive narratives. It would have been difficult to optimize the number of PACs without using human cognition to understand the meanings of words within the context used and to grasp similarities in the meaning among words that can characterize a motive narrative. The feedback loop between the text mining and ETC layers indicates where the cognition process injected decisions about combining or expanding the selected PACs.

3.3. Machine Learning

The ETC workflow discussed above helped the author to assign class labels to each record based only on the motive narrative. The ML process reused the text mining procedures to extract word features from the event summary narrative, which is a different field in the dataset. Unlike the motive field, the event summary narrative described the actions taken and the outcomes of attacks. The ML methods used features extracted from the event summary narrative to predict the PACs, which were the labels derived from the motive narrative. Therefore, good predictive performance would validate the association of the assigned PACs with the event narrative.

Figure 3 shows the workflow of the data, text mining, ML, and output layers. The bag-of-words (BOW) model simplified the representation of each narrative as an array of unordered words derived from the corpus (Aggarwal 2015). Hence, the input to the ML model is a data table where each row contains a word array (features) with an associated class label assigned by the ETC.

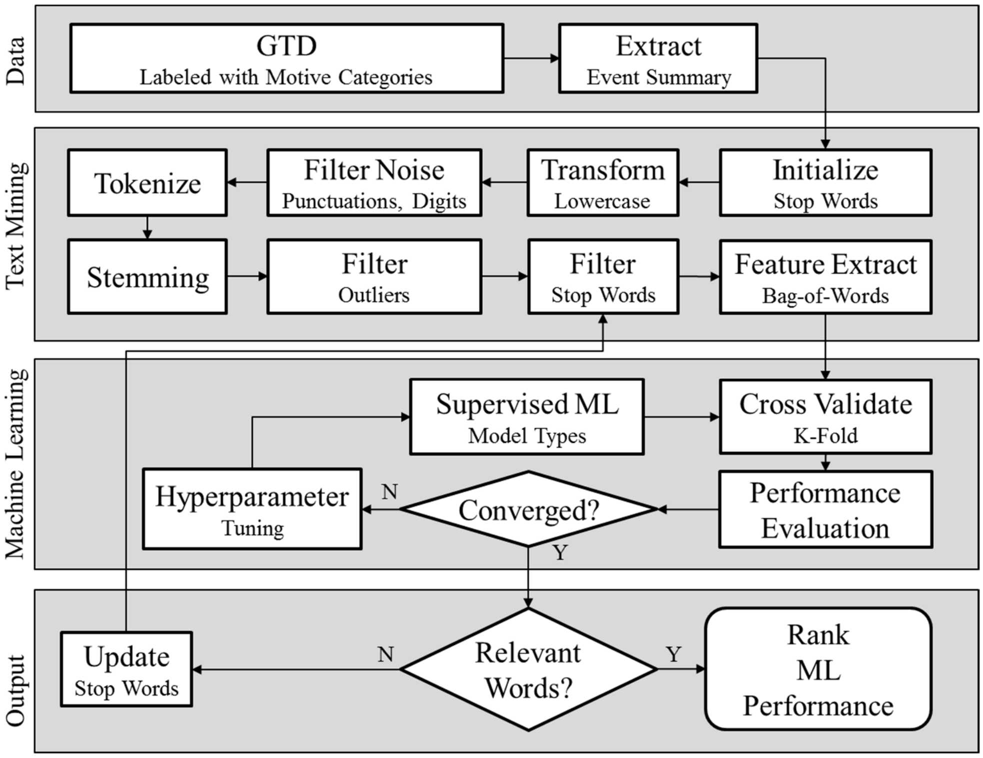

Figure 3. Algorithm for feature extraction by NLP and ML.

The cross-validation (K-fold) procedure cyclically partitioned the narratives into ten subsets. With each iteration, the model output was a prediction error rate for one of the partitions after training on the union of the remaining partitions. The advantage of the k-fold cross validation approach over the traditional single-split of the data into a training and a testing set is that all the data served as test data, which minimized bias and maximized model generalization. The performance evaluation procedure reported the average value of each test metric across all train-test cycles. This cyclical method also helped the author to balance model complexity by adjusting the hyperparameters to tradeoff overfitting and underfitting to the training data. That is, the cross-validation helps to balance model complexity for generalization on unseen data (Géron 2017).

The hyperparameter tuning procedure iteratively adjusted the various model parameters within a defined range until the selected performance metric showed no further improvement. The training procedures used the one-versus-rest (OvR) method to transform the multiclass classification problem into binary classification problem with respect to each class (Géron 2017). This strategy enabled the use of models that natively implement only binary classification. Hence, it became easier to use identical metrics to compare the performance of all models. Table 1 summarizes the fundamental theory of operation for each of the 11 ML models used to validate the classification of PACs derived by machine-aided human decision. The last column of Table 1 includes references that expands on the details of the mathematical formulation and practical implementation of each ML model.

Table 1. Machine learning models applied in the procedural framework.

Model	Description	Reference
Logistic Regression (LR)	Fits the data to a logistic function of the linear combination of attributes to estimate the probability of a binary class.	Aggarwal (2015), Géron (2017)
Support Vector Machine (SVM)	Finds a hyperplane in multidimensional feature space that maximally separates the classes.	Aggarwal (2015), Géron (2017)
Stochastic Gradient Descent (SGD)	Fits a linear multivariate function to the data by randomly selecting data instances to calculate parameter updates that minimize a selected loss function.	Géron (2017)
Decision Tree (DT)	Grows a logic tree by recursively splitting nodes to maximize the purity of child or leaf nodes.	Géron (2017) Hastie et al. (2016)
Random Forest (RF)	Grows many shallow and partial decision trees by randomly selecting a subset of attributes and data subset to split nodes, and then uses majority vote to predict the class.	Aggarwal (2015), Breiman (2001)
AdaBoost (ADB)	Sequentially build shallow decision trees (stumps) that improve on the prediction errors of previous trees, and then uses majority vote to predict the class.	Hastie et al. (2016)
Multi-layer Perceptron (MLP)	A feed-forward and fully connected artificial neural network that learns a function with one or more inner layers of neurons.	Géron (2017) Veen and Leijnen (2016)
Naïve Bayes (NB)	Uses Bayes probability theory to predict a class given the observed set of features, and assuming that they are independent.	Aggarwal (2015), James et al. (2013)
k-Nearest Neighbors (kNN)	Predicts a class based on the majority vote of its k-nearest neighbors in feature space.	Aggarwal (2015), James et al. (2013)
Gradient Boosting (GB)	Sequentially build improved models to predict the errors or residuals of previous models.	Natekin and Knoll (2013), Aggarwal (2015),
Extreme Gradient Boosting (XGB)	A highly configurable version of gradient boosting that incorporates regularization.	Chen and Guestrin (2016), Feng et al. (2020)

The performance evaluation procedure used five scoring metrics: classification accuracy (CA), precision (PR), recall (RC), F1-score, and ROC-AUC score. Each metric used some combination of the classification result based on their the true positive (TP), true-negative (TN), false-positive (FP), and false-negative (FN) rates. The CA was the proportion of correct predictions. The PR, a measure of specificity, was the proportion of positive predictions that were correct: PR = TP/(TP + FP). The RC, a measure of sensitivity, was the proportion of positive predictions recalled from the true positives: RC = TP/(TP + FN). F1 was the harmonic mean of sensitivity and specificity: F1 = TP/(TP + α) where α = (FN + FP)/2.

Each metric is useful for interpreting the performance of a classifier and they complement each other. CA is simple to compute and provides an intuitive sense for comparing the accuracy of each method. However, CA does not effectively characterize the intelligence of a classifier because it can produce a high score for data with high class imbalance if the model predicts the dominant class each time. CA also does not provide any information about the types of errors that a classifier made to help the model designer make decisions to tradeoff FP and FN errors. Therefore, the model designer uses PR, RC, and F1 scores to evaluate the impact of classifier decision boundary thresholds. Although the CA, PR, RC, and F1 scores are simple and useful, they do not provide an unbiased result under high class imbalance. Therefore, the performance evaluation was based on the ROC-AUC score, which integrated TP as a function of FP across a range of sensitivity thresholds to avoid bias from class imbalance (Fawcett 2006).

The ROC-AUC stands for "area under the curve" of the receiver operating characteristic (ROC), where the "curve" is a two-dimensional plot of the TP rate against the FP rate, both as a function of the class membership probability threshold. For brevity, the remainder of this article refers to the ROC-AUC score as simply the AUC score. The AUC provides a relative comparison of classifier performance across a range of decision thresholds. Users can better interpret CA than AUC when evaluating the confidence of a class prediction. However, the ROC can guide threshold selection to maximize CA, PR, RC, and F1 performance. The analysis also included the scores of a null classifier to serve as a baseline for comparing relative performance. The null classifier is a no-skill model that predicts the dominant class every time.

4. Results

The next two subsections present the results of the ETC and the ML classification to validate the accuracy of associating the summary narratives with PACs.

4.1. Topic Classification

The author terminated the ETC procedure of Figure 2 after labeling 8400 of the 9460 records where the remaining dominant words appeared in less than 0.1% of the unassigned records. Table 2 summarizes the six PACs identified and provides a definition for each. Table 2 also list the number of motive narratives or documents (D) classified into each PAC and the number of word features (F) identified for the collection of documents within each PAC. The last column of Table 2 shows a list of some notable word features and their frequency of occurrence across documents within that PAC. The Corpus View helped the author to both identify and verify the context of each keyword and to reconcile them with the assigned PAC. Figure 4 shows the regional makeup of each PAC by the number of attacks assigned that aim category. Note that this chart shows the aim categories for only those terrorist events with motives reported, not all terrorist events. Table 3 complements Figure 4 by providing details of the world region proportions that make up each PAC.

Table 2. Results of the ETC.

PAC	Definition	Corpus	Motive Narrative Keywords and Their Document Frequency
Protest	Expression of ire without necessarily demanding a change.	F: 2620 D: 1544	protest (0.242), elect (0.233), resist (0.069), construct (0.064), celebration (0.049), oppose (0.046), drug (0.037), democrat (0.036), demonstrate (0.035), policy (0.029), republican (0.027), trial (0.026), show (0.024), conflict (0.024), express (0.019), trade (0.014), pipeline (0.013), leftist (0.010), amnesty (0.010), instillation (0.005), miner (0.004), unhappy (0.004), animal (0.003), attention (0.001)

Table 2. *Cont.*

PAC	Definition	Corpus	Motive Narrative Keywords and Their Document Frequency
Retaliate	Reaction to an action or situation.	F: 3498 D: 2956	retaliate (0.335), spy (0.090), NPA (0.078), revenge (0.071), punish (0.048), extort (0.045), avenge (0.035), murder (0.033), crime (0.028), cooperate (0.022), corrupt (0.020), collaborate (0.019), anarchist (0.017), retribution (0.017), traitor (0.014), respond (0.011), mistreat (0.010), reprise (0.010), critic (0.009), sympathetic (0.007), clash (0.006), anniversary (0.004), blame (0.004), react (0.004), defy (0.003)
Intimidate	Repel people or actions by instilling fear.	F: 866 D: 781	intimidate (0.800), threaten (0.060), fear (0.036), threat (0.027), prove (0.020), deter (0.019), raid (0.015), away (0.013), discourage (0.010)
Weaken	Weaken the target or create instability.	F: 617 D: 703	destabilize (0.498), weaken (0.296), disrupt (0.156), interest (0.048), destroy (0.038), offense (0.031), undermine (0.014), dampen (0.009), chaos (0.007), distract (0.007), insurgency (0.004),
Force	Force or demand a change or divert attention.	F: 2806 D: 1638	demand (0.173), liberate (0.090), free (0.066), peace (0.040), prevent (0.038), pay (0.027), land (0.023), sharia (0.022), control (0.020), join (0.017), withdraw (0.016), halt (0.013), remove (0.013), cause (0.012), create (0.011), freedom(0.010), obtain (0.010), taken (0.010), close (0.007), kidnap (0.007), vivisection (0.007), supply (0.006), divert (0.006), preserve (0.006), finance (0.005), incite (0.005), indoctrinate (0.005), autonomy (0.005), revenue (0.004), occupy (0.004), overthrow (0.004), release (0.004), captor (0.002), stop (0.002), surrender (0.002), disavow (0.001)
Despise	Expression of contempt for a race, gender identity, religion, or ideology.	F: 1341 D: 778	Islam (0.288), religious (0.109), anti (0.087), Muslim (0.068), Ramadan (0.064), separatist (0.063), white (0.060), Christian (0.044), pro (0.041), Shi'i (0.041), sectarian (0.033), foreign (0.030), Jew (0.028), women (0.019), nationalist (0.017), ideology (0.015), racial (0.014), music (0.010), refuge (0.009), black (0.008), incompetence (0.008), politician (0.006)

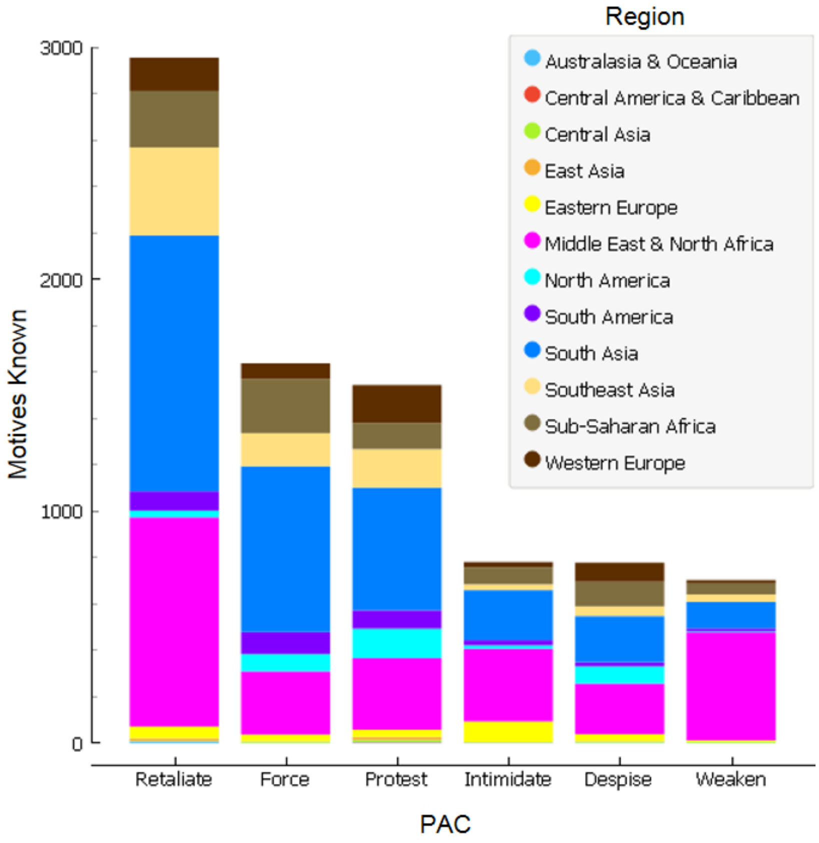

Figure 4. PAC distribution in regions of the world.

Table 3. Proportion of each PAC by region of the world.

World Region	Despise	Force	Intimidate	Protest	Retaliate	Weaken
Australasia and Oceania	0.26%	0.06%	0.26%	0.19%	0.24%	0.14%
Central America and Caribbean	0.00%	0.00%	0.00%	0.13%	0.03%	0.00%
Central Asia	0.51%	0.37%	0.26%	0.58%	0.07%	0.85%
East Asia	0.39%	0.12%	0.13%	0.84%	0.37%	0.00%
Eastern Europe	3.86%	1.71%	11.40%	2.01%	1.73%	0.71%
Middle East and North Africa	28.02%	16.67%	40.08%	20.01%	30.51%	66.15%
North America	9.51%	4.52%	1.92%	8.23%	0.98%	0.71%
South America	2.31%	5.80%	2.82%	5.12%	2.81%	1.85%
South Asia	25.58%	43.65%	27.66%	34.26%	37.31%	16.22%
Southeast Asia	5.40%	8.73%	3.20%	10.75%	12.86%	4.55%
Sub-Saharan Africa	13.62%	14.29%	9.35%	7.25%	8.22%	6.83%
Western Europe	10.54%	4.09%	2.94%	10.62%	4.87%	1.99%

4.2. Machine Learning

Table 4 summarizes the performance metric for each model, ordered by their average AUC score from the OvR binary classification. Table 4 also shows the hyperparameter settings as [*a*:*b*, *x*] where *a*:*b* represents the range of values evaluated using a combined grid-search and manual search method, and *x* is the tuned value of the final classifier. The training and tuning procedures used 10-fold cross validation and stratified sampling. Common hyperparameters for several of the models were the learning rate (L), loss function (LF), regularization (R) setting, and optimizer algorithm (OA). The hyperparameters for tree-based methods were the number of decision trees (N) and the minimum number of samples to retain in the leaves. For kNN, N was the number of nearest neighbors.

Table 4. Average scores for class predictions using OvR classification and hyperparameter settings.

Model	AUC	CA	F1	PR	RC	Hyperparameter Settings
XGB	0.900	0.905	0.895	0.900	0.905	γ:[0:1,0], Max Depth: [2:8,6], Min Child Weight: [0:2,1], R:[0:10,1], w:[0:2,1], L:[0:1,0.2]
RF	0.893	0.890	0.866	0.897	0.890	Trees (*N*): [20:100,60], Attributes/Split: [2:6,5], Min Subset: [2:10,5]
LR	0.887	0.898	0.893	0.892	0.898	R: [L1: L2, L2], C:[0.1:10, 5]
MLP	0.867	0.889	0.885	0.883	0.889	Hidden Nodes: [2:200,100], Activation: ReLu, OA: Adam (α:10^{-4})
NB	0.853	0.696	0.744	0.868	0.696	No parameters to tune.
GB	0.850	0.893	0.875	0.891	0.893	LF: LR, Trees (*N*): [20: 200,100], L: 0.2, Min Samples Leaf: [1:5,1]
SGD	0.841	0.891	0.874	0.882	0.891	LF: (LR, ε:1), R: E.Net (α:10^{-5}, 0.15), L: IVS (η_0:10^{-2}, t:0.25)
kNN	0.769	0.856	0.819	0.857	0.856	*N*: [3:100,30], Distance (Euclidean, Weights: Uniform)
ADB	0.768	0.890	0.885	0.883	0.890	Trees (*N*): [20:200,50], LF: Linear, OA: SAMME.R, LR: 1.0
DT	0.735	0.866	0.838	0.842	0.866	Max Depth: [5:20,10], Min Samples Leaf (*N*): [5:200,90], Min Subset: [5:20,5]
SVM	0.624	0.262	0.236	0.785	0.262	Kernel: Sigmoid, R (C:0.2, ε:1.0)
Null	0.499	0.833	0.761	0.703	0.833	No parameters to tune.

The XGB algorithm provided the best performance across all metrics. Table 5 summarizes the performance metrics of the XGB algorithm for each PAC in the OvR classification. Figure 5 provides a visual validation of the differences in AUC among the classifications. The gray background highlights the area that is equivalent to the AUC. The figure shows the ROC mean (green curve) and variations (error bars) for the 10 cross-validation results of the XGB algorithm when predicting the PAC presence for "Weaken" and "Despise" which represents the best and worst AUC scores, respectively. The TP rate of the ROC curve for

"Weaken" increases more rapidly than that of the ROC curve for "Despise" which has a less sharp curve. For reference, Figure 5 includes the performance of the Null model, which is the straight orange line.

Table 5. XGB scores for predicting each class using OvR classification.

Category	AUC	CA	F1	PR	RC
Protest	0.906	0.904	0.896	0.901	0.904
Retaliate	0.921	0.865	0.861	0.868	0.865
Intimidate	0.914	0.940	0.933	0.935	0.940
Weaken	0.927	0.943	0.940	0.938	0.943
Force	0.870	0.860	0.839	0.853	0.860
Despise	0.864	0.920	0.899	0.906	0.920
Mean	0.900	0.905	0.895	0.900	0.905
STD	0.027	0.036	0.039	0.035	0.036
CV	0.030	0.040	0.044	0.038	0.040

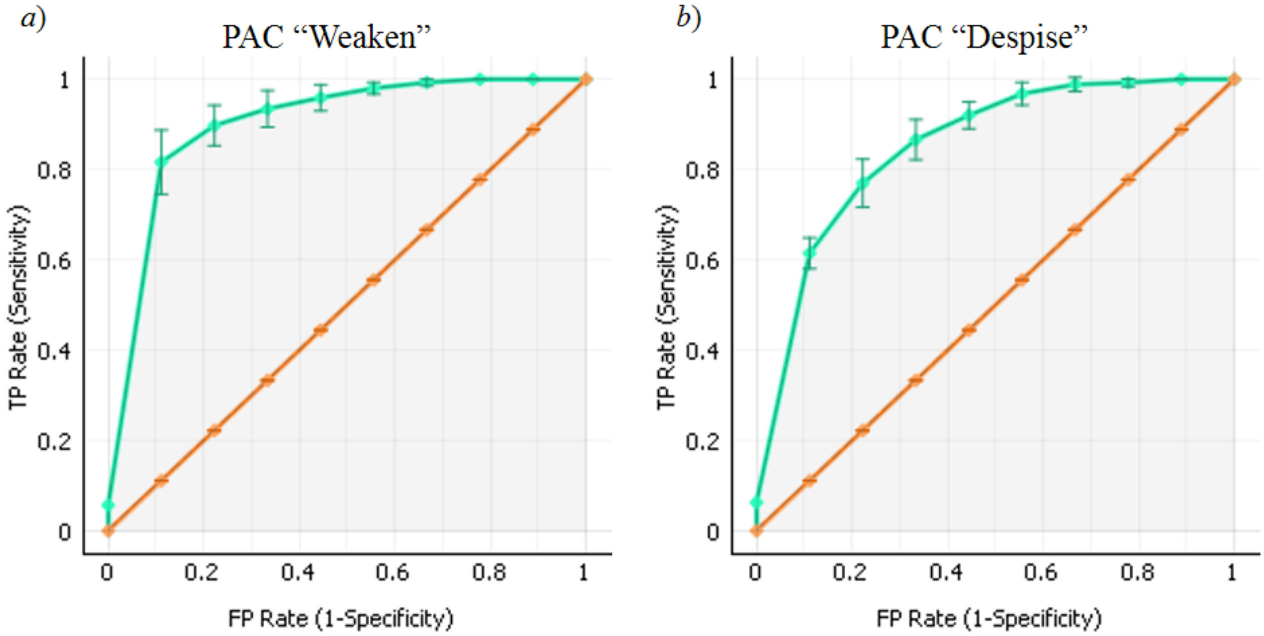

Figure 5. Comparison of ROC curves for the prediction of (**a**) PAC "Weaken" and (**b**) PAC "Despise".

5. Discussion

The AUC values ranged from 0.624 for the SVM model to 0.900 for the XGB model, with the Null model achieving only 0.499. The ROC uses the entire dataset to compute probability rates (TP and FP) between 0 and 1 as a function of decision thresholds applied to the output of each classifier. Therefore, the AUC is inherently a probabilistic score. This means that decision makers can interpret differences in AUC values on a probability scale when selecting models. The six PACs accurately classify the summary narratives of the events. That is, the RF, LR, MLP, and GB models provided similarly satisfactory performance. The null model provided the worst AUC performance as expected.

Although in a statistical sense, the six models of XGB, RF, LR, MLP, NB, and GB provided similar performance, the consistent albeit slight performance edge of XGB across all scores is worth an interpretation. As explained in the introduction, an important tenant of ML is that no single model type can best represent all datasets (Aggarwal 2015). For example, SVM does best when there is a clear separation of classes across hyperplanes, whereas k-NN does well when classes form clusters in feature space (Géron 2017). XGB incorporates a lot of randomness as it sequentially builds an ensemble of models to reduce the errors of previous models (Natekin and Knoll 2013). Therefore, XGB is more likely to find class associations in datasets that lack clear hyperplane separation or clustering because it can randomly discover them. The original reference described the algorithm as "especially appropriate for mining less than clean data" (Friedman 2001).

The CA performance of the XGB classification ranged from 0.860 to 0.943 for each PAC (Table 5). The table summarizes the mean, standard deviation (STD), and coefficient of variation (CV) for each score. This result indicated that even though XGB made more prediction errors for some classes than others, the performance spread as measured by CV was less than 5%, suggesting a consistent performance across all classes. The variations in CA for each target class was partially due to imbalance in the number of training examples across the dataset (Table 2). For example, the PAC of "Retaliate" had 2956 documents whereas "Weaken" had 76% fewer documents to serve as examples in the training. Another reason for the differences were inherent variations due to randomness in the cross-validation processes.

The main advantage of the ETC over the available topic modeling techniques was the accommodation of human cognition into the process to yield a meaningful set of PACs. Hence, the workflow enabled a machine-aided human decision process. Consequently, the process was naturally slower than a fully automated statistical method. Hence, the tradeoff for more meaningful topic classification was speed. For example, using the same computing resources, LDA took a few minutes whereas the ETC took an entire day. As discussed previously, LDA did not provide meaningful results whereas ETC did. In this application, the ETC proved to be effective because of the relatively small size of the dataset, and the similarities of contextual keywords across many documents.

A potential limitation of the approach is that the motives and summary narratives of the GTD may have reporting biases. However, it is unlikely that any potential bias would have been systematic and consistent across all reports. Machine learning algorithms continue to evolve in their capabilities. The framework discussed is sufficiently general to include additional model types when they become available, but it is possible that some approaches may require a modification to the framework for compatibility. General users may be more familiar with CA as an evaluation metric. Hence, the framework may require advanced users to evaluate other performance metrics such as F1 and AUC to guide more informed decision-making and interpretation.

One interpretation of Figure 4 was that among known motives, aims to "retaliate" were nearly four-times more revealed by terrorist responsibility claims than aims to "weaken" the target. This suggests that terrorists were more likely to claim credit for retaliations than to advertise aims to weaken or distract their enemies. Another interesting observation is that PACs were more generally known for attacks in the Middle East, North Africa, and South Asia than for other regions of the world.

A final observation is that the six PACs represent two general need categories. The first need is to instigate change by weakening the target, forcing a situation, or intimidating the opposition. The second need is to express feelings through despise, protest, or retaliation. These findings, with the help of the two AI methods (NLP and ML), add to the diversity of knowledge about terrorist behaviors.

6. Conclusions

A comprehensive review of the research on terrorism revealed a dominant assumption that terrorists attack civilians to maximize political ends. However, "revealed preference" by the reported motives of terrorist attacks points to a more diverse set of aims. This research applied two artificial intelligence (AI) techniques to help the author distill perpetrator aims from a comprehensive record of terrorist attacks across the world. The method helped the author to identify six aim categories, which were to despise, protest, retaliate, weaken, force, and intimidate.

The performance of the machine learning (ML) techniques validated that the six perpetrator aim categories (PACs), derived by machine-aided human decision, accurately classified the data. The method first applied natural language processing (NLP) techniques to invent an empirical topic classification workflow that extracted text features from the available motive narratives of terrorist events. The method then applied 11 distinct types of ML models to text features extracted from the summary narrative of each event, which was a different field from the motive narrative that was not involved in the model training. The six best-performing ML models predicted PACs with accuracies ranging from 86% to 94% and corresponding AUC scores ranging from 86% to 93%. The Extreme Gradient Boosting model provided the best predictive performance with AUC, classification accuracy, and F1 scores of 0.90, 0.91, and 0.90, respectively. These values provide a probabilistic confidence level of the classification results relative to the outcome from other models evaluated. The level of confidence can inform resource allocation decisions to address an aim identified from terrorist event reports.

A common strategy in counterterrorism is to diminish incentives, which must be known beforehand. Hence, the intelligence community can view the PACs as an incentive structure to help identify and customize strategies for deterrence. To further the research agenda in counterterrorism, future work will examine if patterns of terrorist activities can be associated with one of the six aim categories. Future work will also assess possible reasons for the dominance of certain PACs in different regions of the world.

Funding: This research received no external funding. The APC was funded by the College of Business, North Dakota State University.

Data Availability Statement: The data that support the findings of this study are openly available in the START Consortium at https://www.start.umd.edu/data-tools/global-terrorism-database-gtd. Accessed on 13 April 2020.

Conflicts of Interest: The author declares that there is no conflict of interest.

References

Abrahms, Max. 2008. What Terrorists Really Want: Terrorist Motives and Counterterrorism Strategy. *International Security* 32: 78–105. [CrossRef]

Abrahms, Max, Nicholas Beauchamp, and Joseph Mroszczyk. 2017. What Terrorist Leaders Want: A Content Analysis of Terrorist Propaganda Videos. *Studies in Conflict & Terrorism* 40: 899–916.

Aggarwal, Charu C. 2015. *Data Mining*. New York: Springer International Publishing, p. 734.

Araújo, Rafaella de C. R., Magdalena Bobowik, Roosevelt Vilar, James H. Liu, Homero Gil de Zuñiga, Larissa Kus-Harbord, Nadezhda Lebedeva, and Valdiney V. Gouveia. 2020. Human values and ideological beliefs as predictors of attitudes toward immigrants across 20 countries: The country-level moderating role of threat. *European Journal of Social Psychology* 50: 534–46. [CrossRef]

Arce, Daniel G. 2019. On the human consequences of terrorism. *Public Choice* 178: 371–96. [CrossRef]

Balestrini, Pierre Philippe. 2021. Counterterrorism Evaluation and Citizens: More Than about Policing? *The Social Sciences* 10: 298. [CrossRef]

Bassetti, Thomas, Raul Caruso, and Friederich Schneider. 2018. The tree of political violence: A GMERT analysis. *Empirical Economics* 54: 839–50. [CrossRef]

Bilal, Muhammad, Abdullah Gani, Muhammad Ikram Ullah Lali, Mohsen Marjani, and Nadia Malik. 2019. Social Profiling: A Review, Taxonomy, and Challenges. *Cyberpsychology, Behavior, and Social Networking* 22: 433–50. [CrossRef]

Björgo, Tore. 1993. Terrorist violence against immigrants and refugees in Scandinavia: Patterns and motives. In *Racist Violence in Europe*. Edited by Björgo Tore and Rob Witte. London: Palgrave Macmillan, pp. 29–45.

Breiman, Leo. 2001. Random forests. *Machine Learning* 45: 5–32. [CrossRef]

Burke, Paul. 2021. Al-Qaeda. In *Global Jihadist Terrorism*. Edited by P. Burke, D. Elnakhala and S. Miller. Cheltenham: Edward Elgar Publishing, p. 352.

Burnap, Peter, and Matthew Leighton Williams. 2014. Hate speech, machine classification and statistical modelling of information flows on Twitter: Interpretation and communication for policy decision making. Paper presented at the Internet, Policy & Politics, Oxford, UK, September 26.

Canetti, Daphna, Joshua Gubler, and Thomas Zeitzoff. 2021. Motives Don't Matter? Motive Attribution and Counterterrorism Policy. *Political Psychology* 42: 483–99. [CrossRef]

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: 441–52. [CrossRef]

Cassese, Antonio. 2006. The Multifaceted Criminal Notion of Terrorism in International Law. *Journal of International Criminal Justice* 4: 933–58. [CrossRef]

Chen, Tianqi, and Carlos Guestrin. 2016. XGBoost: A Scalable Tree Boosting System. Paper presented at the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, August 13–17.

Coaffee, Jon. 2009. Protecting the Urban The Dangers of Planning for Terrorism. *Theory, Culture & Society* 26: 343–55.

Coccia, Mario. 2018. A Theory of General Causes of Terrorism: High Population Growth, Income Inequality and Relative Deprivation. *Social Science Research Network* 2: 26. [CrossRef]

Cottee, Simon. 2021. Incel (E)motives: Resentment, Shame and Revenge. *Studies in Conflict & Terrorism* 44: 93–114.

Cottee, Simon, and Keith J. Hayward. 2011. Terrorist (E)motives: The Existential Attractions of Terrorism. *Studies in Conflict & Terrorism* 34: 963–86.

Das, Priyanka, and Asit Kumar Das. 2019. Graph-based clustering of extracted paraphrases for labelling crime reports. *Knowledge Based Systems* 179: 55–76. [CrossRef]

Enders, Walter, and Xuejuan Su. 2007. Rational Terrorists and Optimal Network Structure. *Journal of Conflict Resolution* 51: 33–57. [CrossRef]

Fawcett, Tom. 2006. An introduction to ROC analysis. *Pattern Recognition Letters* 27: 861–74. [CrossRef]

Feng, Yi, Dujuan Wang, Yunqiang Yin, Zhiwu Li, and Zhineng Hu. 2020. An XGBoost-Based Casualty Prediction Method for Terrorist Attacks. *Complex & Intelligent Systems* 6: 721–40.

Friedman, Jerome H. 2001. Greedy function approximation: A gradient boosting machine. *Annals of Statistics* 29: 1189–232. [CrossRef]

Géron, Aurélien. 2017. *Hands-On Machine Learning with Scikit-Learn and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems*, 2nd ed. Sebastopol: O'Reilly Media, p. 856.

Guo, Diansheng, Ke Liao, and Michael Morgan. 2007. Visualizing patterns in a global terrorism incident database. *Environment and Planning B: Planning and Design* 34: 767–84. [CrossRef]

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: 133. [CrossRef]

Hastie, Trevor, Robert Tibshirani, and Jerome Friedman. 2016. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd ed. New York: Springer, p. 767.

Höflinger, Tim. 2021. Modern terrorism: Motives of individual terrorists or the strategies of terrorist groups? *Global Change, Peace & Security* 33: 77–83.

Huamaní, Enrique Lee, Alva Mantari, 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: 562–570. [CrossRef]

Huff, Connor, and Joshua D. Kertzer. 2018. How the Public Defines Terrorism. *American Journal of Political Science* 62: 55–71. [CrossRef]

James, Gareth, Daniela Witten, Trevor Hastie, and Robert Tibshirani. 2013. *An Introduction to Statistical Learning with Applications in R*. New York: Springer, vol. 112.

Jaspersen, Johannes G., and Gilberto Montibeller. 2020. On the Learning Patterns and Adaptive Behavior of Terrorist Organizations. *European Journal of Operational Research* 282: 221–34. [CrossRef]

Jones, Karen Ssparck, and Peter Willett, eds. 1997. *Readings in Information Retrieval*. Burlington: Morgan Kaufmann, p. 589.

Khalifa, Nour Eldeen Mahmoud, Mohamed Hamed N. Taha, Sarah Hamed N. Taha, and Aboul Ella Hassanien. 2019. Statistical Insights and Association Mining for Terrorist Attacks in Egypt. In *International Conference on Advanced Machine Learning Technologies and Applications*. Cham: Springer, pp. 291–300.

Krieger, Tim, and Daniel Meierrieks. 2011. What causes terrorism. *Public Choice* 147: 3–27. [CrossRef]

Kurtulus, Ersun N. 2017. Terrorism and fear: Do terrorists really want to scare? *Critical Studies on Terrorism* 10: 501–22. [CrossRef]

Kydd, Andrew H., and Barbara F. Walter. 2006. The Strategies of Terrorism. *International Security* 31: 49–79. [CrossRef]

Lane, Hobson, Cole Howard, and Hannes Max Hapke. 2019. *Natural Language Processing in Action: Understanding, Analyzing, and Generating Text with Python*. Shelter Island: Manning Publications Co.

Luo, Lanjun, and Chao Qi. 2021. An analysis of the crucial indicators impacting the risk of terrorist attacks: A predictive perspective. *Safety Science* 144: 105442. [CrossRef]

Mashechkin, Igor V., Mikhail Petrovskiy, Dmitry V. Tsarev, and Maxim N. Chikunov. 2019. Machine Learning Methods for Detecting and Monitoring Extremist Information on the Internet. *Programming and Computer Software* 45: 99–115. [CrossRef]

Masuku, Mfundo Mandla, Victor H. Mlambo, and Bhekani J. Ngwenya. 2021. The Critical Analyses of Propaganda of the Terrorism Deed. *Technium Social Sciences Journal* 25: 619–29.

Maszka, John. 2018. The Perils of Deduction: Limitations of Theoretical Models in Terrorism Studies. *Systems Research and Behavioral Science* 35: 884–907. [CrossRef]

Mishra, Namrata, Shrabanee Swagatika, and Debabrata Singh. 2020. An Intelligent Framework for Analysing Terrorism Actions Using Cloud. In *New Paradigm in Decision Science and Management. Advances in Intelligent Systems and Computing*. Edited by Srikanta Patnaik, Andrew W. H. Ip, Madjid Tavana and Vipul Jain. Singapore: Springer, vol. 1005, pp. 225–35.

Monahan, Tom, and Robin Maria Valeri. 2018. Terrorism and Fear. In *Terrorism in America*, 1st ed. Edited by Robin Maria Valeri and Kevin Borgeson. New York: Routledge, p. 256.

Morley, Barry, and Gavin D. Leslie. 2007. Terrorist bombings: Motives, methods and patterns of injuries. *Australasian Emergency Nursing Journal* 10: 5–12. [CrossRef]

Natekin, Alexey, and Alois Knoll. 2013. Gradient Boosting Machines, a Tutorial. *Frontiers in Neurorobotics* 7: 21. [CrossRef]

Padmaja, V. R., S. Lakshmi Narayana, and C. H. Divakar. 2018. Probabilistic Topic Modeling and its Variants: A Survey. *International Journal of Advanced Research in Computer Science* 9: 173–77.

Reich, Walter. 1990. Understanding terrorist behavior: The limits and opportunities of psychological inquiry. In *Origins of Terrorism: Psychologies, Ideologies, Theologies, States of Mind*. Edited by Walter Reich. Cambridge: Cambridge University Press, Washington, DC: Woodrow Wilson International Center for Scholars, pp. 261–79.

Rigterink, Anouk S. 2021. The Wane of Command: Evidence on Drone Strikes and Control within Terrorist Organizations. *American Political Science Review* 115: 31–50. [CrossRef]

Rothenberger, Liane Phil, and Kathrin M.A. Müller. 2015. Categorizing terrorist entities listed by the European Union according to terrorist groups' underlying motives. *Conflict & Communication* 14: 1–14.

START. 2020. University of Maryland. Available online: https://www.start.umd.edu/data-tools/global-terrorism-database-gtd (accessed on 13 April 2020).

Tamborini, Ron, Lindsay Hahn, Melinda Aley, Sujay Prabhu, Joshua Baldwin, Neha Sethi, Eric Novotny, Brian Klebig, and Matthias Hofer. 2020. The Impact of Terrorist Attack News on Moral Intuitions. *Communication Studies* 71: 511–27. [CrossRef]

Treistman, Jeffrey. 2021. Social Exclusion and Political Violence: Multilevel Analysis of the Justification of Terrorism. *Studies in Conflict & Terrorism*, 1–24. [CrossRef]

Uddin, M. Irfan, Nazir Zada, Furqan Aziz, Yousaf Saeed, Asim Zeb, Syed Atif Ali Shah, Mahmoud Ahmad Al-Khasawneh, and Marwan Mahmoud. 2020. Prediction of Future Terrorist Activities Using Deep Neural Networks. *Complexity* 2020: 1373087. [CrossRef]

Van Um, Eric. 2011. Discussing Concepts of Terrorist Rationality: Implications for Counterterrorism Policy. *Defence and Peace Economics* 22: 161–79. [CrossRef]

Veen, Fjodor Van, and Stefan Leijnen. 2016. The Neural Network Zoo. The Asimov Institute for Artificial Creativity & Constraint. September 14. Available online: https://www.asimovinstitute.org/neural-network-zoo (accessed on 3 August 2020).

Wong, Kevin, Geoffrey Walton, and Gavin Bailey. 2021. Using information science to enhance educational preventing violent extremism programs. *Journal of the Association for Information Science and Technology* 72: 362–76. [CrossRef]