An Integrated Machine Learning Approach To Studying Terrorism

Andi Peng

Advised by Dr. Brian Scassellati,
Professor of Computer Science, Cognitive Science, and
Mechanical Engineering

Submitted to the faculty of Cognitive Science in partial fulfillment of the requirements for the degree of Bachelor of Science

Yale University
April 20, 2018

Abstract

This project investigates an integrated machine learning approach for classification and analysis of global terrorist activity. In this project, we aim to make the following three contributions: 1) exploration of supervised machine learning approaches as a novel technique in the study of terrorist activity; 2) development of a model that classifies historical events in the Global Terrorism Database (GTD) that, at present, have yet to be attributed to a responsible party; and 3) release of a new dataset, QFactors_Terrorism, that integrates event-specific features derived from the GTD with population-level demographic data from open sources like the World Bank and United Nations. Using this new dataset, a random forest model was trained that classifies the actor responsible for an identified incident with up to 68% accuracy. This project makes no claim on the ability to forecast or predict future terrorist activity—rather, it is intended to highlight the importance of a machine learning approach that, when integrated with domain-area expertise, can augment study of complex social issues.

I. Intro

Terrorist attacks are widespread, leading to social destruction and political instability across many nations. Terrorism is defined by the United Nations as “any action with a political goal that is intended to cause death or serious bodily harm to civilians [1]. In 2017, 22,487 events were observed globally, causing over 18,000 direct fatalities [2]. There exists conflicting evidence regarding the exact factors that lead to the deployment of terrorism, and it is likely that these factors change over time in response to key political events and social zeitgeists. Moreover, not only are the factors that cause terrorists to take up arms difficult to identify, attributing the attack in the aftermath to its responsible party is also difficult [3]. The lack of detailed knowledge regarding widespread patterns of terrorist behavior and the prevalence of labor-intensive methods of studying terrorism have proved challenging for individuals who work in the contemporary security space.

Traditionally, studies on terrorism have attempted to study group behavior through a combination of qualitative (case study) and quantitative (regression analysis) methods. For example, a typical analysis of a committed terrorist event in the United Kingdom may include on-the-ground interviews of civilians impacted by the attack combined with linear regression analysis of manually-identified factors, such as weapon used or number of civilians harmed, in identifying features that contribute to the proper identification of the perpetrator. A different analysis may include retroactively filtering through information received from intelligence signals, such as attempting to identify unusual individual behaviors or interrogating detainees *ex-post* for information, in an attempt to attribute the event. Such methods are extremely labor-intensive, often requiring hundreds of analysts, and the results criticized for being ungeneralizable beyond the specific group and/or event studied [4].

This project aims to provide a novel approach to studying terrorism—one that integrates supervised machine learning techniques with terrorism specific domain knowledge to extract macro-level conclusions about the pattern of terrorist behavior. A novel dataset, QFactors_Terrorism, was developed using data compiled from the GTD, World Bank, United Nations, and other open data sources to study population-level demographic features in attributing terrorist events that were previously difficult to study through conventional methods. Through analysis of events using both the unaltered GTD and integrated QFactors_Terrorism datasets, five supervised machine learning models (Gaussian Naïve Bayes, Linear Discriminant Analysis, *k*-Nearest Neighbors Clustering, Decision Tree, and Random Forest) were built and evaluated on their performances in attributing the group responsible for an identified terrorist event. We observe an increase from 26% to 68% in classification accuracy from random forest models trained by the original vs. integrated datasets, suggesting that an integrated machine learning approach combined with domain-area expertise show promising results for studying social complex phenomenon, especially when information is rare or incomplete.

II. Background

a. A Political Science Approach

Terrorist attacks are not a new phenomenon, but the robust theoretical study of terrorism through quantitative methods is. While the public media tends to depict terrorism as a new cultural occurrence beginning with the al-Qaeda attacks on September 11 and continuing through the Islamic State's activity in Iraq and Syria today, the reality is that terrorism has its roots in early resistance and political movements stemming back hundreds of years [5]. On the one hand, the top-line statistics highlight an improvement in the levels of global terrorism over that timeframe. On the other, continued intensification of terrorist events, especially in the past 20 years in specific countries impacting transnational populations, is a cause for serious concern [6].

This may be due to the fluid nature of modern terrorist activity. According to researchers like Wilkinson and Stewart (1987) [7] and Rice (1988) [8], the state of the international order since the end of the Cold War has made engaging in conventional wars, such as with traditional means like tanks and armies, extremely costly. Moreover, technological advancements and the spread of information have disseminated successful terrorist tactics, such as suicide bombings, with incredible ease. This has led to the strategic balance of power currently favoring the use of terror by non-state actors as an unconventional means of engaging with rivals, especially within certain regions of the world [9]. As a result, we've seen a dramatic increase in both the deployment of terrorism as a specific tool as well as the number of academics studying the phenomenon when compared to that in the past.

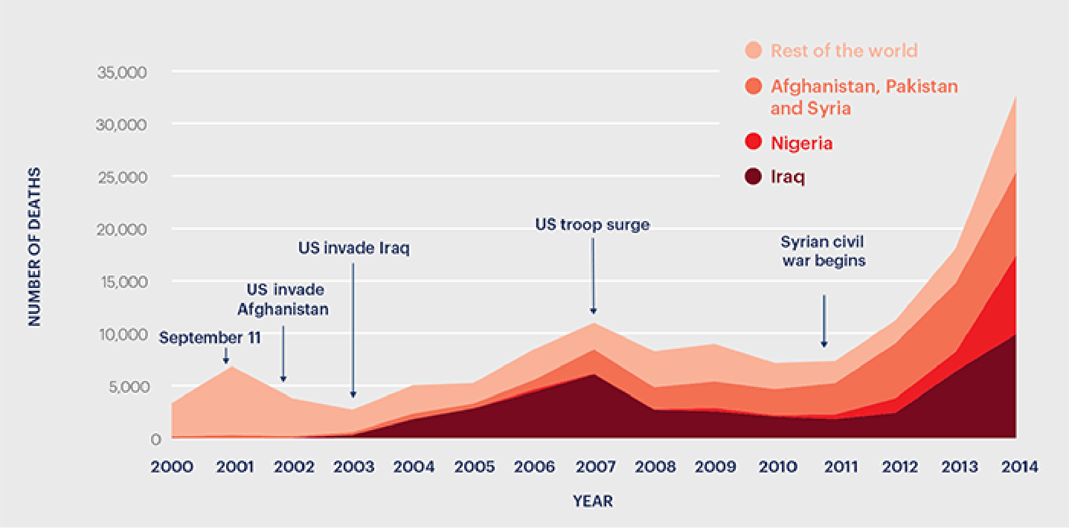


Figure 1: Deaths from terrorism from 2000-2014. The number of people who have died from terrorist activity has increased ninefold since the year 2000 and spike around salient political events [10].

Traditionally, there have been many theoretical approaches to the study of terrorism, some conflicting with others. In the first, *instrumental*, approach, the act of terrorism is studied as a deliberate and rational choice made by a political actor to achieve a goal in response to various external stimuli, such as government policies or social oppression [11]. In the second, *organizational*, approach, the prevalence of violent attacks are hypothesized to be the result of a terrorist organization's struggle for "survival" rather than for ideological motives, often in a competitive environment [12]. The organization responds to existential pressures by providing its members incentives to remain active in the group. In the third, *strategic communication*, approach, terrorist attacks are utilized to spread a public message so that pressure can be placed on a state actor [13]. Thus, a terrorist organization's main metric at evaluating the success of an attack is by the attention that it receives. In the fourth, *economic*, approach, terrorism is theorized to be the result of a lack of economic opportunity [14]. As a result, terrorist organizations provide stability and employment as incentives for members. In line with these approaches, the following factors have been hypothesized as contributing to the spread of terrorism:

Economic Factors

The most popular theory among scholars is that terrorism is rooted in economic deprivation. Although human civilization has, over time, created and refined institutional structures to reduce the level of conflict over limited resources, intra-population fighting remains a perennial feature of society [15]. The few studies that have explored the far-reaching consequences of poverty in weak or failed states find that the poorer the state, the more likely they are to experience revolution. These academics argue for a “greed” narrative, suggesting that people seek to overthrow states because they don’t have physical resources and lack economic prosperity [16]. A variety of studies have linked poverty to terrorism through a variety of factors such as social inequality, low GDP, and low literacy or education levels. Other sources included other factors such as population density, unemployment rates, and inflation [17].

A different economic argument, that of “grievances”, also exists. The “grievances” narrative argues that on a macro level, perceptions of scarcity caused by poverty gaps are generated when there is a discrepancy between what individuals think they deserve and what they actual receive through the economic (distributive) process. In other words, people not only fight when they *are* poor, they also fight because they *feel* poor. This is supported by neuroeconomic literature. Collier in 2004 found that countries with abundant natural resources are more prone to violent conflict than those without because of the perception of inequality generated between the haves and the have-nots. This position is predicated on the supposition that when economic, social and political power differentials exist between heterogeneous groups whether ethnic, linguistic, cultural, religious or any other categorization, the outbreak of conflict motivated by grievances can be predicted extremely accurately [18]. Such a perspective acknowledges that both psychological constraints and environmental instruments combine to produce decision-making factors that influence how combatants choose to engage in violence.

Political Factors

Another highly-cited theory of terrorism suggests that government repression and political instability are also key drivers of terrorism. Samuel Huntington famously theorized in 1996 that clashes between civilizations may result in violence [19]. When groups exhibit different identities (such as race or ethnicity), this may lead to more conflict either between different groups within a nation or between different national groups organized along civilizational lines depending on political perceptions. Such a world view eliminates moral considerations regarding violence and strengthens a group's organizational cohesion, making terrorism less costly and more effective [20]. Key features that have been linked to terrorist behavior as a result of these ingroup-outgroup delineations include imigration and refugee levels, ethnic fractionalization, and religious differences within societies [21].

Furthermore, while it's debated as to which systems of governance are better able to prevent or respond to terrorism, it's been demonstrated that political repression may be linked to terrorist behavior [22]. A series of case studies conducted in 2006 on terrorism in authoritarian states show that the political exclusion and repression of Islamist movements have contributed to the adoption of terrorist methods in some cases [23]. For example, the two leading figures of al-Qaeda, Osama Bin Laden and Ayman al Zawahiri, were citizens of states ruled by repressive regimes, Saudi Arabia and Egypt, respectively. Al-Zawahiri was one of the leaders of al-Jihad (the Egyptian group that assassinated Sadat in 1981) and was instrumental in drawing the organization into international activity by formally merging with al-Qaeda in 1998. It is argued that both were driven to take up arms because they lacked political freedom and stability. Thus, features such as the measure of civil rights and institutionally granted freedoms may also contribute to terrorism.

Because of these conflicting and highly variable approaches to studying terrorism, we have yet to understand which factors are ultimately most important in studying the overall patterns of terrorist groups. Which groups are motivated more by social grievances and ethnic discrimination? Which groups are responding to a lack of economic opportunity? Which groups are instead protesting unfavorable government policies and reduced civil liberties? Until we have a greater and more consistent understanding of how these factors all interact in driving terrorism within specific groups and regions, it will be difficult to attribute future terrorist events to their responsible perpetrators.

b. A Machine Learning Approach

Arthur Samuel once described machine learning (ML) as a field that "gives computers the ability to learn without being explicitly programmed" [24]. Although also not a new field, ML research has experienced a research boom in the past several years due to the availability of high-quality labeled data generated by technology companies and increases in computing power, opening up new markets and opportunities for public impact in critical focal areas such as public health, energy, and national security. In recent years, machines' successes at performing automated tasks using ML have spurred advancements in specific application spaces such as image recognition for the diagnosis of diseases and classification of fake news.

This approach—learning from data—contrasts with the older "expert system" approach in which programmers sit down with human domain experts to learn the rules and criteria used to make decisions [25]. An expert system aims to emulate the principles used by human experts, whereas machine learning relies on statistical methods to find a decision procedure that works well in practice. Through such an approach, the machine is often able to find patterns that are both predictable and not immediately unobservable to a human analyst. Supervised learning, the technique applied in this project, analyzes a training dataset and produces an inferred function, which can be used for mapping new examples for classification of unseen data.

$$Y = f(x)$$

Supervised machine learning is best understood as approximating a target function (f) that maps input variables (X) to an output variable (Y). This is done by providing a training dataset with both the predictive X variables (features) paired with their expected Y outcomes, and allowing an algorithm to train a model using that information. Then, performance of the model is evaluated on data not yet seen and adjusted accordingly. Pedro Domingos summarizes this concept as ‘Learning = Representation + Evaluation + Optimization’ [26].

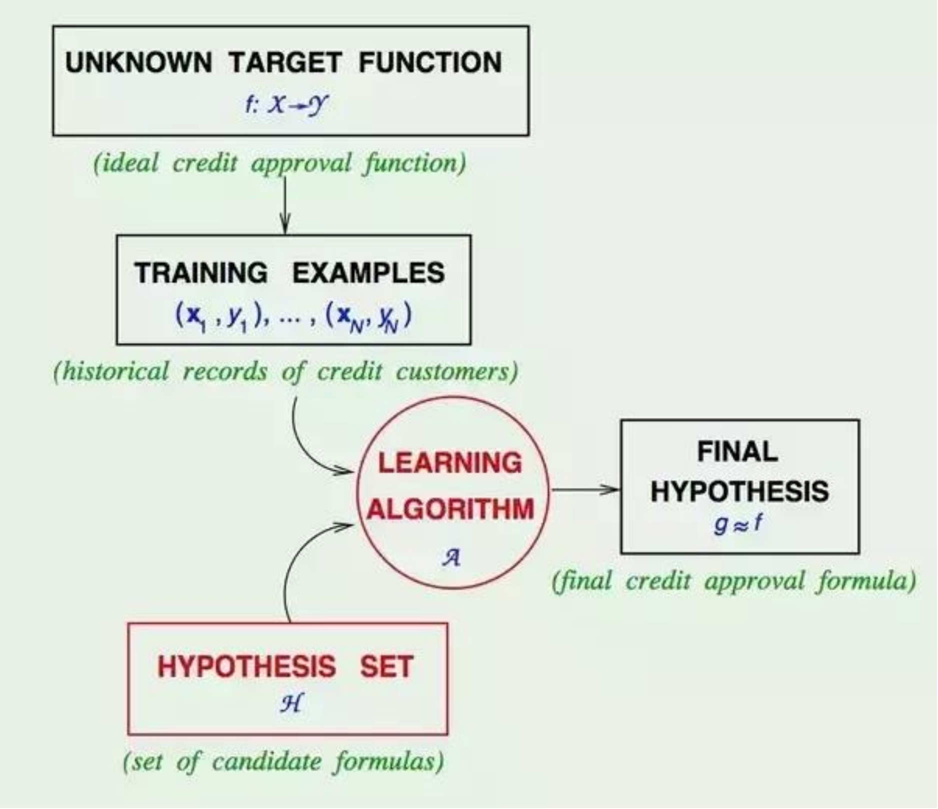
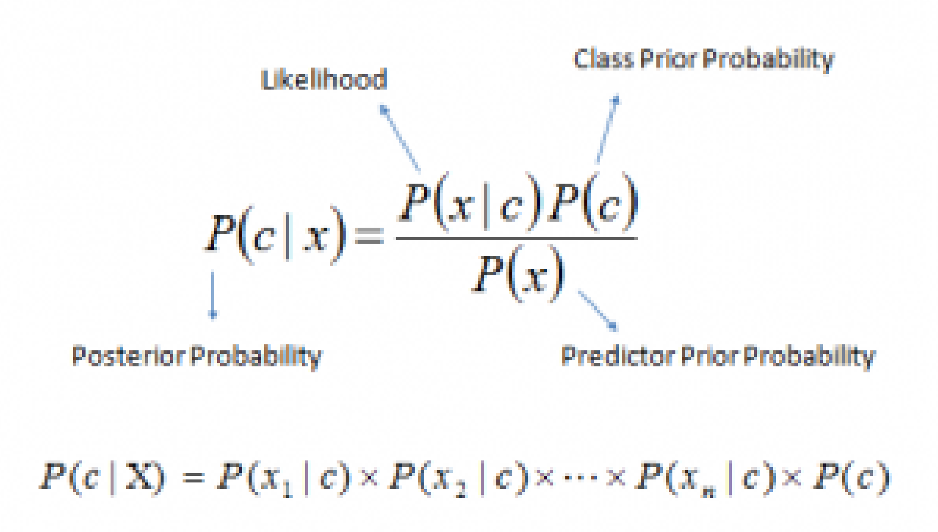

UNKNOWN TARGET FUNCTION
$f: \mathcal{X} \to \mathcal{Y}$
(ideal credit approval function)
TRAINING EXAMPLES
$(\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_N, y_N)$
(historical records of credit customers)
LEARNING ALGORITHM
$\mathcal{A}$
FINAL HYPOTHESIS
$g \approx f$
(final credit approval formula)
HYPOTHESIS SET
$\mathcal{H}$
(set of candidate formulas)

Figure 2: Training of a supervised learning algorithm [27].

There exist many supervised machine learning algorithms that perform classification tasks. In this project, we explore the following five models in classifying terrorist behavior. These models were chosen for the following three reasons: 1) ease of use; 2) robust online documentation; and 3) clearly understood tradeoffs.

Naïve Bayes (NB)

One of the simplest supervised algorithms is a Naïve Bayes classifier. Bayes' heorem provides a method to calculate the probability of a hypothesis given our prior knowledge. A NB classifier builds on this by also assuming that the presence of a particular feature in a class is unrelated to the presence of any other feature.



$$P(c \mid x) = \frac{P(x \mid c)P(c)}{P(x)}$$

$$P(c \mid X) = P(x_1 \mid c) \times P(x_2 \mid c) \times \cdots \times P(x_n \mid c) \times P(c)$$

Figure 3: Naïve Bayes classifier algorithm [28].

This equation is then used to calculate the posterior probability for each class. The class with the highest posterior probability is the outcome of prediction. For example, consider the task of classifying whether al-Qaeda or the Maoists committed a terrorist attack. Training data that is given to us may include attributes describing al-Qaeda events as bombings impacting many civilians while Maoists events as stabbings impacting individual citizens. The NB classifier will, instead of characterizing relationships between these attributes and attempting to weight them together, consider each of these attributes separately when classifying a new instance of an event seen.

NB is relatively simple and intuitive to understand. Furthermore, it is easily trained with both small and large datasets and its runtime is relatively fast. When the assumption of independence holds, a NB classifier performs better than other models like logistic regression with less training data [28]. However, true independence is rarely seen in real-world applications [29].

Linear Discriminant Analysis (LDA)

LDA is also based off of Bayes' Theorem. However, instead of estimating P(c|x) directly, estimates of its distribution as a multivariate normal distribution are computed. Mathematically, the algorithm trains by searching the data for a linear combination of predictors (features) that best separates different classes.

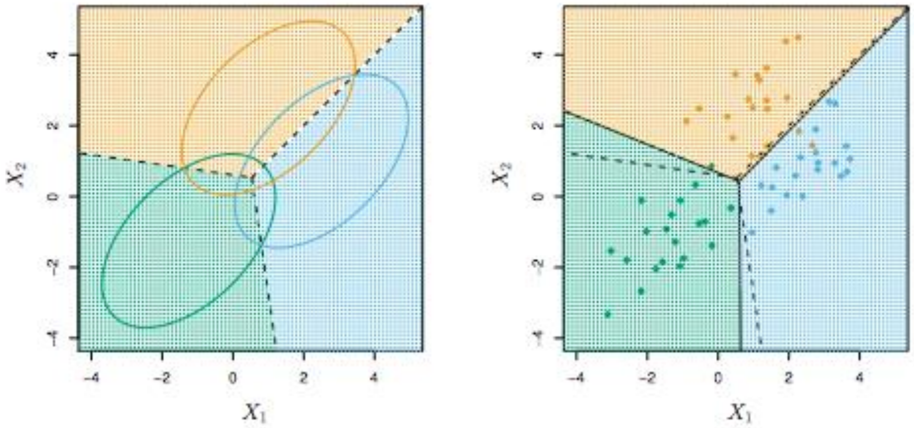


Figure 4: Linear decision boundaries between classes in LDA [30].

When provided a test observation, the predicted class is then classified by estimating the fraction of training samples that fall within those linear decision boundaries. LDA will always output an explicit solution and is computationally convenient due to its low-dimensionality, but suffers from the assumption that linear separability can be achieved in all classifications.

k-Nearest Neighbors (k-NN) Clustering

k-NN is another algorithm commonly used for supervised classification problems. First introduced in 1951, the algorithm aims to identify homogeneous subgroups such that observations in the same group (clusters) are more similar to each other than others [31]. Each data points' *k*-closest neighbors are found by calculating Euclidean or Hamming distance and grouped into clusters. The *k*-closest data points are then analyzed to determine which class label is the most common among the set. The most common class is then classified to the data point being tested. For *k*-NN classification, an input is classified by a majority vote of its neighbors. That is, the algorithm obtains the classification of its *k* neighbors and outputs the class that represents a majority of the *k* neighbors.

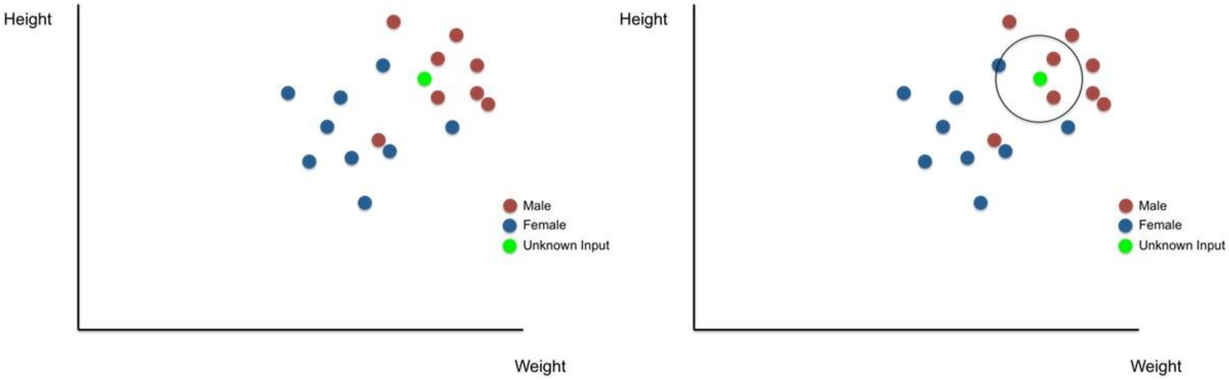


Figure 5: Example *k*-NN clustering for classification of gender (male, female, unknown) based on height and weight [32].

k-NN is a non-parametric algorithm, meaning it makes no assumption regarding the probability distribution of its inputs, and is thus more robust than parametric algorithms which must assume properties about input data. It is also intuitively extremely easy to understand. However, the tradeoff comes with more computational time required as all computation is done during testing, instead of training [33]. Furthermore, normalization is required if one class appears more often than another, for the classification of an output will also be more biased towards that class (since it is more likely to be neighbors with the input).

Decision Tree

Decision tree classifiers organize a series of test questions and conditions in a tree structure. The goal is to create a model that predicts the value of a target variable by learning simple decision rules inferred from the data features. In a tree, the root and internal nodes contain attribute test conditions to separate nodes that have different characteristics. Inputs are entered at the top and traverse down the tree, following the appropriate branches as the data gets bucketed into smaller and smaller sets. A class is assigned once the input has reached a terminal node.

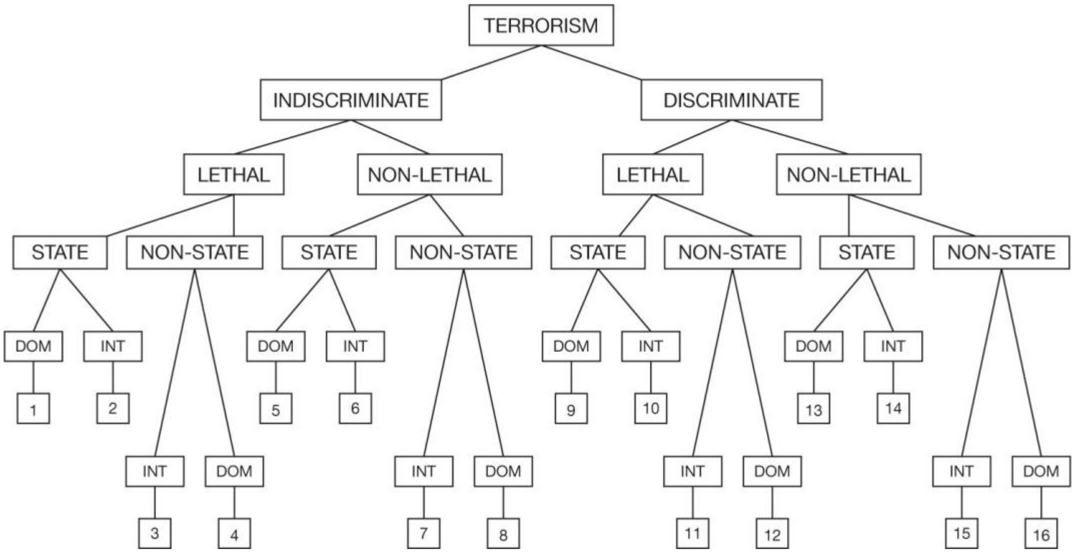


Figure 6: Example decision tree as illustrated by Kaplan [34].

Decisions trees can be easily visualized, which allows for easy comprehension and traceback of decisions made. Furthermore, they have the ability to handle continuous as well as discrete data. However, both higher classification error rates are observed when the training set is small in comparison with the number of classes (too many terminal nodes compared to branches, thus causing overfitting) [35].

Random Forest (RF)

An RF is simply a collection of decision trees. The random forest starts with training many different decision trees and combining them into an ensemble, the “forest”. Then, when classifying a new unknown data point, each decision tree will test the observation and vote on which class it believes the observation to be. By majority vote, the random forest will output the most likely classification.

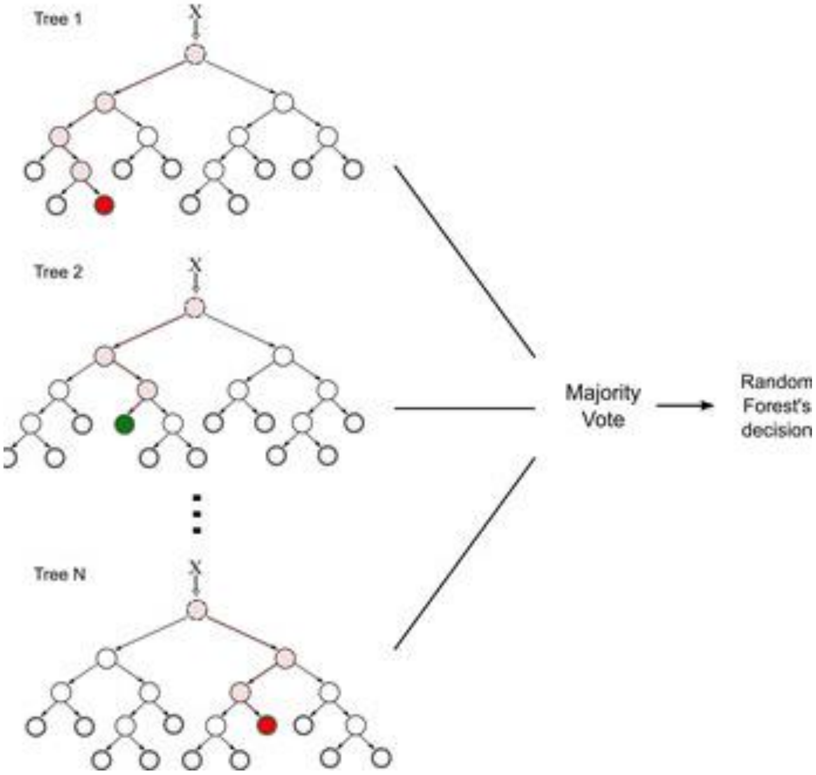


Figure 7: Random forest [36].

An RF can be thought of as an ensemble approach that is similar to nearest neighbor predictor [37]. Ensembles are a divide-and-conquer approach used to improve performance. The main principle behind ensemble methods is that a group of "weak learners" can come together to form a "strong learner". RFs correct a decision tree's tendency to overfit by constructing a multitude by which to aggregate classification from. However, some of the interpretability of a single tree is lost, and computational complexity also increases exponentially.

Limitations

There also exist a variety of limitations that plague all the models assessed in this project. The performance of any supervised learning model is entirely dependent upon the representation of the data it receives [38]. For example, if researchers wish to develop a method by which to predict the likelihood of an individual defaulting on a loan, they would train a model using various factors, or *features*, such as age, credit history, employment, etc. that they believe to be most useful in predicting the outcome variable, which in this case would be the probability of default. Then, they take a predetermined amount of inputs from a training dataset and train a model that may predict the original dataset correctly with, perhaps, 99% accuracy. However, oftentimes the model, when tested on a new (unseen) dataset, fails to perform nearly as well. Therein lies the fundamental tradeoff that plagues researchers: machine learning models do not often *generalize* well when faced with new data because the model was *overfitted* to the training data. This concept is encapsulated as the *bias-variance* tradeoff: the problem of simultaneously minimizing two sources of error (over and underfitting) that prevent supervised learning algorithms from generalizing beyond their training set [39].

To limit overfitting, several techniques, such as feature selection or regularization, are utilized in this project. The most common technique, *cross-validation*, is a resampling technique often seen as a gold standard. In cross-validation, the initial training data is used to generate multiple mini train-test splits. These splits are then used to tune the model before evaluation. For example, a standard *k*-fold cross-validation partitions the data into *k* subsets, called folds. Then, the machine learning model is iteratively trained on *k*-1 folds while using the remaining fold as the test set (called the "holdout fold"). In this way, parameters utilized by the model can be tuned with only the original training set, allowing the test set to remain unseen until evaluation.

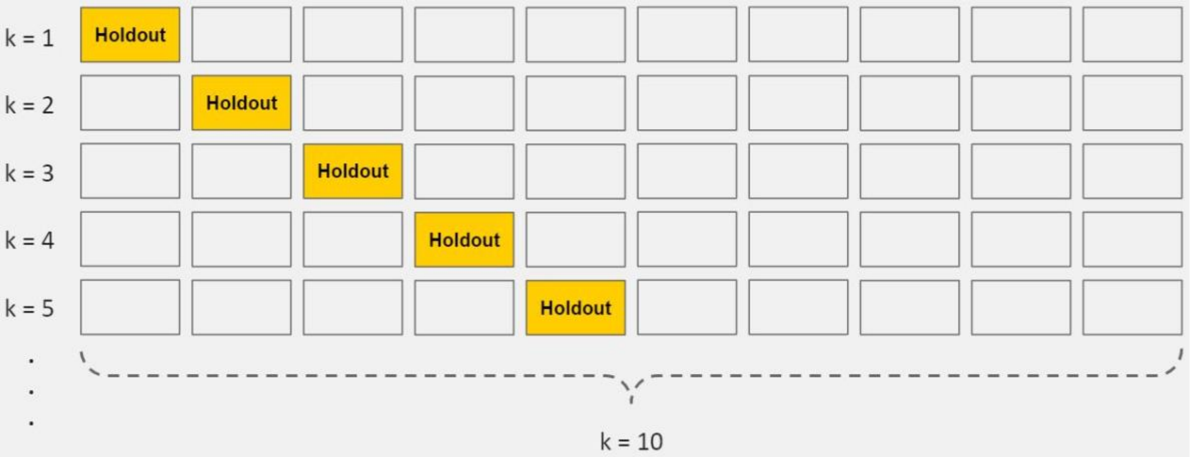


Figure 8: *k*-fold cross-validation where *k* = 10 [40].

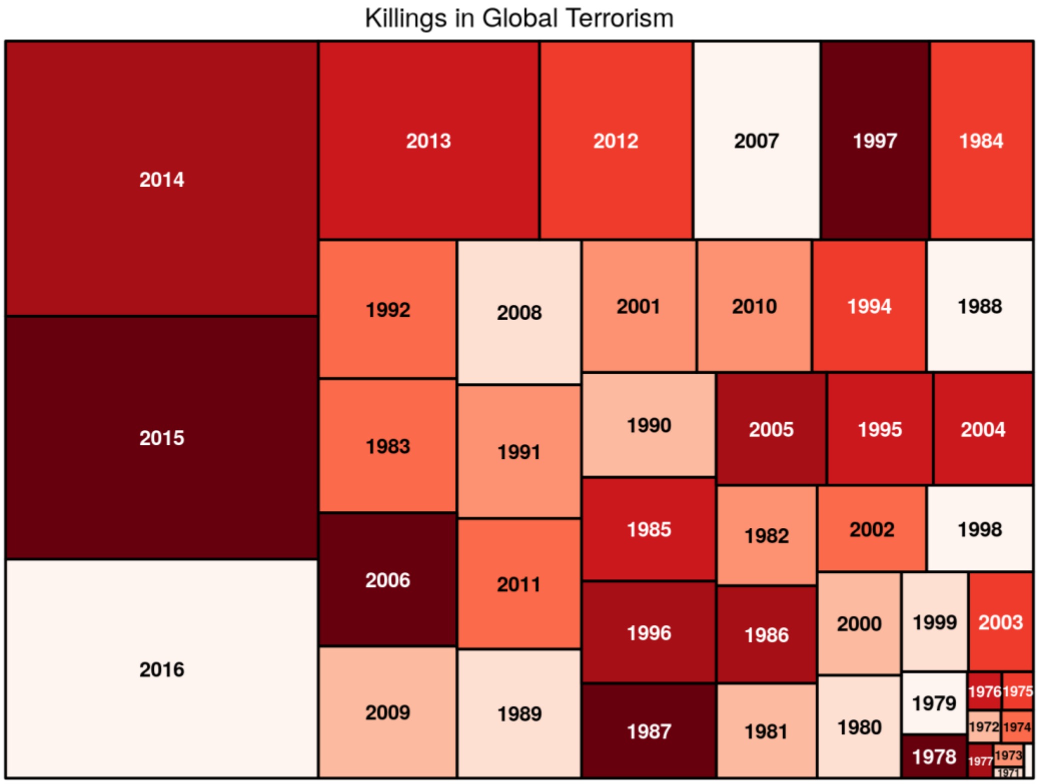
III. Development of an Integrated Dataset

a. Global Terrorism Database (GTD)

For this project, the most recent 2017 release of the Global Terrorism Database (GTD), a dataset collected and collated by the National Consortium for the Study of Terrorism and Responses to Terrorism (START), a Department of Homeland Security Centre of Excellence led by the University of Maryland, was utilized as the original dataset. The GTD is considered to be the most comprehensive dataset on terrorist activity globally and has now codified over 170,000 terrorist incidents from 1970-2016 [41]. For each GTD incident listed, information is available on the details associated with the specific event in question such as date and location of the incident, the weapon(s) used and nature of the target, the number of casualties, and—when identifiable—the group or individual responsible. It is important to note that the GTD does not contain population-level data beyond the specified incident.

Statistical information contained in the GTD is based on reports from a variety of open media sources, such as newspapers and UN reports. According to researchers who maintain the database, information is not added to the GTD "unless and until [they] have determined the sources are credible". See the GTD Codebook for more details on data collection methodology, definitions, and coding schema.

Table 1: Inspection of event distributions in the GTD from 1970-2016. The number of fatalities have steadily increased in the past 20 years, confirming the deadliness of terrorist events.



The following adjustments were made to impute missing data. When possible, unknown values are recoded to maintain consistency with the GTD's original coding:

- `extended`: NaN values recoded as 0 (No)
- `success`: NaN values recoded as 0 (No)
- `suicide`: NaN values recoded as 0 (No)
- `attacktype1`: NaN values recoded as 9 (Unknown)
- `targtype1`: NaN values recoded as 20 (Unknown)
- `subtargtype1`: NaN values recoded as 112 (Unknown)
- `natlty1`: NaN values recoded as 1005 (Unknown)
- `weaptype1`: NaN values recoded as 13 (Unknown)
- `weapsubtype1`: NaN values recoded as 27 (Unknown)
- `nkill`: NaN values imputed from the mean

Table 2: Correlation matrix of identified features of the GTD.
`iyear extended, success, suicide, attacktype1, individual, weaptype1, weapsubtype1, nkill`
were ultimately kept as predictors for the GTD model after examination.

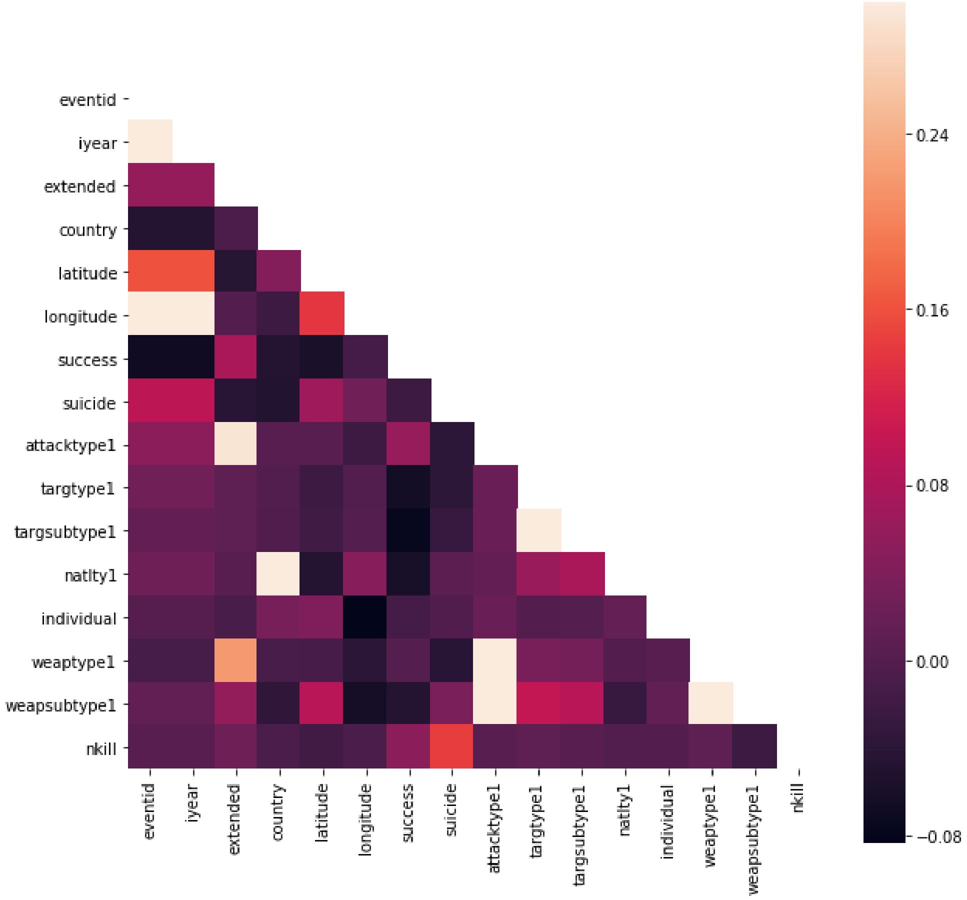
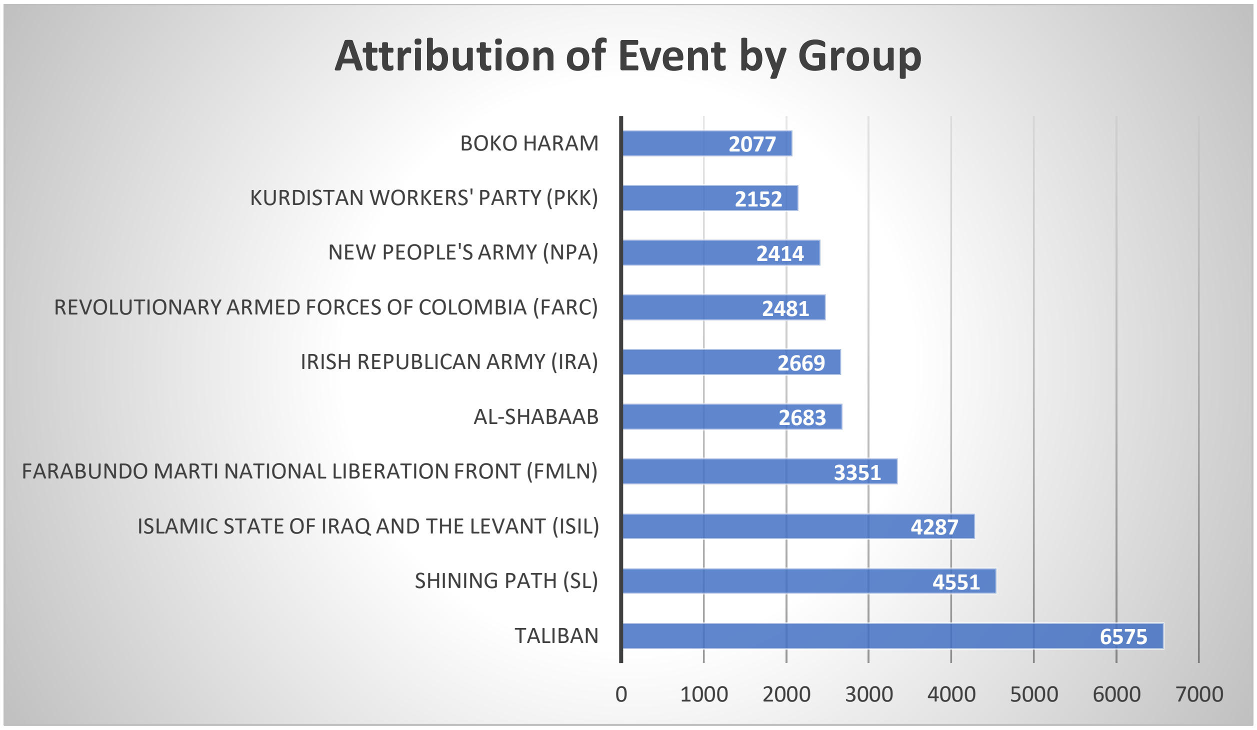


Table 3: Incidents attributed to unique groups in the GTD (3454 in total). Most notably, unknown groups represent the vast majority of recorded incidents, with 78306 observations out of 170350 (45.97%).



b. Development of QFactors_Terrorism

In this project, we propose an integrated machine learning approach to studying terrorism, one that incorporates the expertise of political scientists, is proposed. As highlighted above, there are a combination of economic and political factors that have long been studied as factors contributing to terrorism. Thus, intuition posits the question of whether the attribution of terrorism events to their perpetrators may be studied by datasets encompassing additional population-level features that are not encompassed by the GTD. If these patterns exist and interact with the outbreak of violence, then they could be potentially studied through a machine learning approach.

The following features and datasets were utilized in compilation of QFactors_Terrorism, a new dataset with additional population-level data associated with each incident recorded in the GTD. Due to availability of data, only observations from 1990 onwards were retained from the GTD.

Table 4: Features of QFactors_Terrorism and their original sources.

Feature Information	Dataset	Source
event identifier group responsible event year extended conflict? event country event city event latitude event longitude successful? suicide? attack type (kidnapping, bombing, etc.) target type (civilian, government worker, etc.) target nationality event by individual perpetrator? weapon type (biological, chemical, etc.) number killed	Global Terrorism Database (GTD)	National Consortium for the Study of Terrorism and Responses to Terrorism (START), University of Maryland
education level	Human Development Report (HDR)	United Nations Development Programme (UNDP)
number of immigrants number of refugees	International Migration Stock	United Nations Department of Economic and Social Affairs (UN DESA)
GNI per capita (PPP) life expectancy population density primary school enrollment	World Development Indicators (WDI)	The World Bank
violence by government	Armed Conflict Dataset	Uppsala Conflict Data Program (UCDP) / Peace Research Institute Oslo (PRIO)
political freedom	Human Freedom Index (HFI)	Cato Institute

IV. Models

For this project, the programming language Python was used for both compilation of QFactors_Terrorism as well as development of the models. The following open-source packages were utilized: [pandas, numpy, scikit-learn]. The name of the group responsible for the event was isolated as the outcome variable, with event-specific features (weapon used, number of people killed, etc.) selected for evaluation of the GTD and population-level features (primary school enrollment, poverty ratio, etc.) selected for evaluation of QFactors_Terrorism.

After collecting and creating the two datasets (the GTD and QFactors_Terrorism), five common supervised machine learning algorithms were trained and evaluated on both datasets: Gaussian Naïve Bayes (GNB), Linear Discriminant Analysis (LDA), *k*-Nearest Neighbor Clustering (KNN, where *k* = 5), Decision Tree (CART), Random Forest (RF)

For each dataset, both time independent and dependent models were evaluated for their accuracy, with an additional feature (event year) included in the dependent models. Principal component analysis (PCA) was conducted to reduce the dimensionality of both training sets (from 9 to 4 features with 99.7% variance retained for GTD and from 7 to 3 features with 99.6% variance retained for QFactors_Terrorism). *k*-fold cross validation (*k* = 10) was utilized in sampling from the training sets for model evaluation.

Table 5: Description of training and test sets for all models evaluated.

	Time Independent Model	Time Dependent Model
Training Set	Random 80% of data from 1990-2016	Sub training set: Random 80% of data from 1990-2014 Validation set: Remaining 20% of data from 1990-2014
Test Set	Remaining 20% of data from 1990-2016	All data from 2015-2016

V. Results

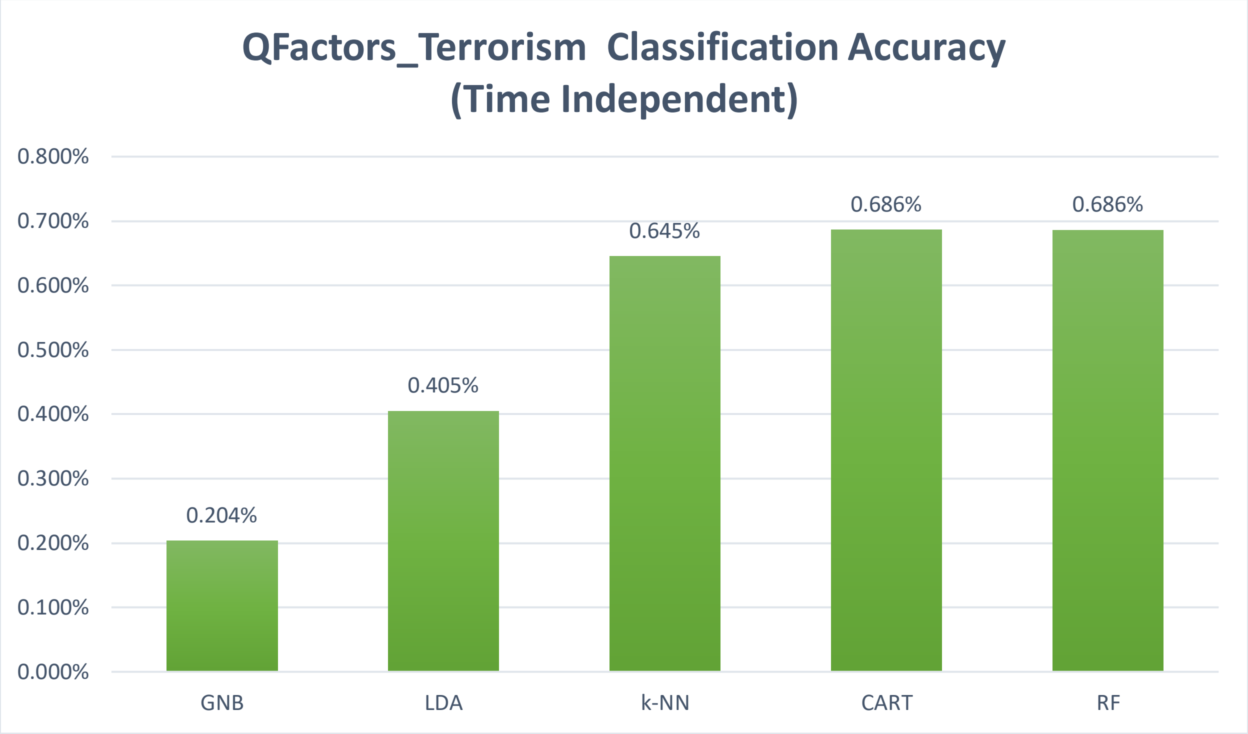
Model Comparison

Three main results were found. First, as expected, neither GTD nor QFactors_Terrorism trained models performed as well on the time dependent tasks when compared to the time independent, suggesting that the ability to "forecast" future events given historical data is limited due to overfitting of past models. This is shown in the 12% loss in accuracy from models trained with the GTD and 10% loss in accuracy from models trained with QFactors_Terrorism. Second, there was a marked improvement in accuracy in classification from models trained on individual_level features from the GTD when compared to population-level features from QFactors_Terrorism, suggesting that there is a link between demographic, social, and economic factors in a country influencing specific terrorist group behavior. Third, while GNB and LDA performed the worst, *k*-NN was comparable to CART and RF in classification accuracy.

Table 6: A comparison of classification accuracy across models:

	Time Independent Model	Time Dependent Model
GTD	GNB: 0.063961% LDA: 0.146788% KNN: 0.211290% CART: 0.259277% RF: 0.262480%	GNB: 0.002762% LDA: 0.088684% KNN: 0.131912% CART: 0.144473% RF: 0.147858%
QFactors_Terrorism	GNB: 0.203594% LDA: 0.404838% KNN: 0.645166% CART: 0.686455% RF: 0.685956%	GNB: 0.210192% LDA: 0.482551% KNN: 0.558761% CART: 0.528250% RF: 0.583483%

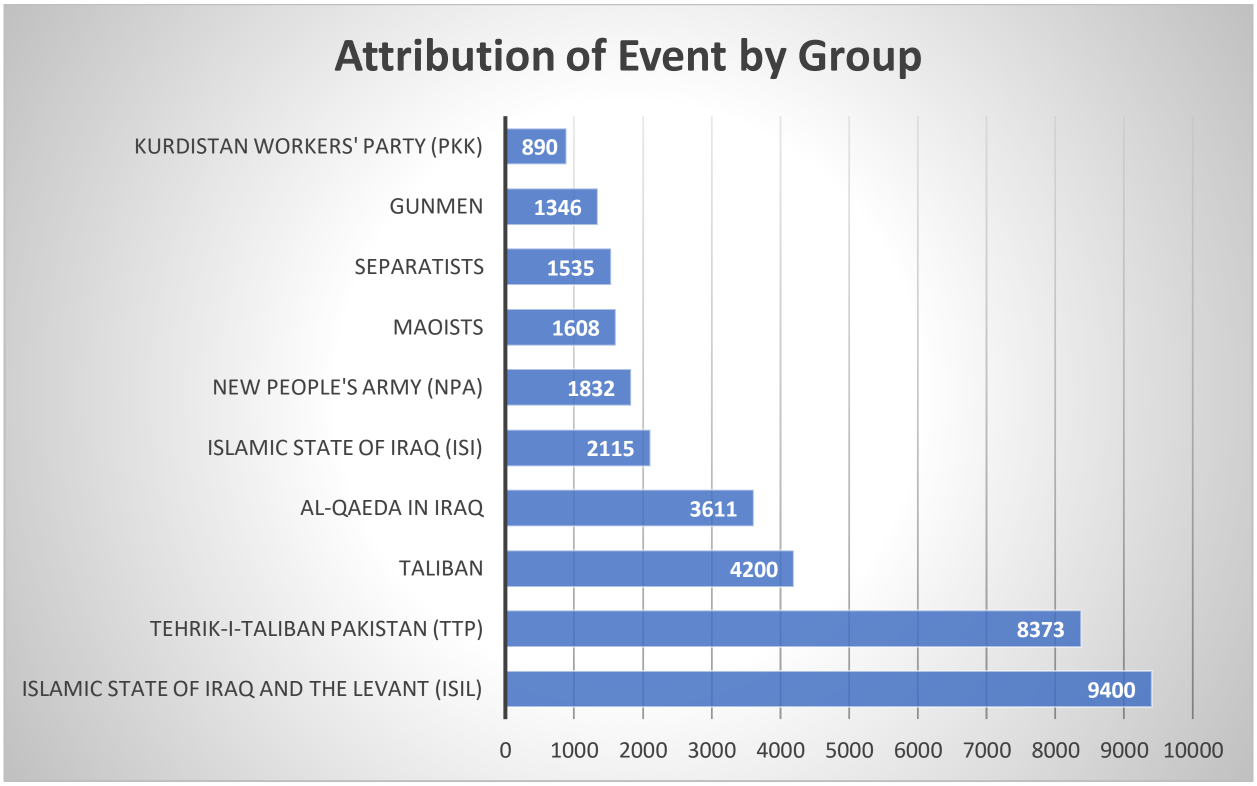
Table 7: Comparison of model accuracy trained from QFactors_Terrorism in evaluating events from the full dataset (time independent).



Prediction of Unknown Groups

We employ the highest performing time independent model, CART (random forest) trained by QFactors_Terrorism to predict the original unknown groups from the full GTD dataset.

Table 8: Incidents attributed to unique groups in the GTD with predictions. The Islamic State of Iraq and the Levant (ISIL) overtakes the Taliban as the most active group.



VI. Conclusion

Discussion

In this work, we wish to highlight how machine learning and other advanced statistical techniques can be integrated with domain-area expertise from the social sciences to explore patterns difficult to study through either perspective alone. Even though we explored only a small subset of the potential models and algorithms possible in attributing group responsibility of terrorist attacks, we demonstrate that the performance of a classification task can be dramatically improved through compilation of a higher quality dataset informed by literature of the field of application.

Although this work proves promising for future study, it is important to emphasize that the predictive power of these models is extremely limited, or in other words, we can make no claim regarding the ability to forecast future terrorist behavior. The inherent statistical problem is that, like many other complex social issues, terrorism is a low-frequency event and every single event can be seen as unique, which means that the risk of low base rate fallacy and over-generalization increases [42]. While immense to the field of political science, the dataset employed in this project is rather small when compared to those in typical machine learning research, where millions if not billions of data points are sometimes fed as the training set. By running both time-independent and -dependent models, we observed overfitting of the model, suggesting that there are still outside features not encompassed by either dataset that hold explanatory power towards the outcome variable.

Moreover, there exist immense social implications, many undesirable, of this type of work if utilized but uninformed decisionmakers. For example, it would not be the intention of this project for the predictions of group attribution of unknown events in the GTD to be used in the retroactive criminal prosecution of those events, regardless how high the statistical accuracy achieved. This project also cannot be taken to make any statement regarding the ethics or fairness of classification in such a manner, especially on issues as sensitive and ever-changing as terrorism.

Considerations for Future Research

To some researchers, the long-term research vision of machine learning is to build high-fidelity predictive models capable of informing them of events that have yet to happen by understanding patterns of the past [44]. However, the lack of available training data continues to prove a challenging, and likely insurmountable, hurdle preventing the generalizability of models beyond their training data. In studying low-probability rare events such as terrorism, machine learning researchers must take care to ensure that claims of high predictive accuracy are couched in such terms, lest they be employed in a potentially damaging fashion.

Lastly, future developers of automated decision-making or -aiding systems must take great care to ensure that there is proper tradeoff between algorithmic accuracy and fairness. When the only training data provided may potentially be skewed in such a manner as to discriminate against certain classes by age, race, or ethnicity—the proper balance must be struck between constraining the classifier to not be overly discriminative with respect to sensitive features while also preserving the power of statistical inference. In other words, fairness is achieved through awareness [45].

Acknowledgements

To:

Brian Scassellati for his invaluable support and mentorship;

Mark Sheskin for his guidance throughout the entire thesis process;

The Yale Computer Science department

for training me to be a technical thinker;

The Yale Jackson Institute for Global Affairs

for training me to be a qualitative thinker;

The Yale Cognitive Science program

for allowing me to be an academic mutt;

and my fellow peers who inspired me to put it all together.

Appendix

For all code, codebooks, and datasets: please refer to: github.com/andipeng

References

[1] “United Nations Terrorism,” *United Nations*, (2017). [Online].

[2] A. Erickson and L. Karklis, “Every 2017 Terrorist Attack, Mapped,” The Washington Post, 18-Jan-2018. [Online].

[3] B. Doosje, S. Zebel, M. Scheermeijer, and P. Mathyi. Attributions of Responsibility for Terrorist Attacks: The Role of Group Membership and Identification. *International Journal of Conflict and Violence,* 1(2):127-141, (2007).

[4] T. Rid and B. Buchanan. Attributing Cyber Attacks. *Journal of Strategic Studies*, 38(1-2):4-37, (2014).

[5] G. Michael. Leaderless Resistance: The New Face of Terrorism. *Defence Studies*, 12(2):257–282, (2012).

[6] Institute for Economics and Peace. *Global Terrorism Index: Measuring and Understanding the Impacts of Terrorism*, (2016). [Online].

[7] P. Wilkinson and A. M. Stewart. *Contemporary research on terrorism*. Aberdeen: Aberdeen University Press, (1989). [Print].

[8] E. Rice. Wars of the third kind: *Conflict in underdeveloped countries*. Berkeley: University of California Press, (1990). [Print].

[9] J. Baylis, S. Smith, and P. Owens. *The globalization of world politics: an introduction to international relations*. Oxford: Oxford University Press, (2017). [Print].

[10] M. Clarke. “Globally, terrorism is on the rise - but little of it occurs in Western countries.” *ABC News*, 17-Nov-2015. [Online].

[11] M. Crenshaw. *Terrorism in context*. University Park, Pennsylvania State University Press, (2007). [Print].

[12] D. Rapoport. *Inside terrorist organizations*. London: Frank Cass, (2001). [Print].

[13] H. Ingram and A. Reed. Lessons from history for counter-terrorism strategic communications. *Terrorism and Counter-Terrorism Studies*, (2016). [Print].

[14] P. Collier. *Breaking the conflict trap: civil war and development policy*. Washington, DC: World Bank, (2008).

[15] S. Pinker. *The better angels of our nature: why violence has declined*. New York: Penguin, (2012).

[16] S. Rice, C. Graff, and C. Pascual. *Confronting poverty weak states and U.S. national security*. Washington, DC: Brookings Institution Press, (2010).

[17] G. Akhmat, K. Zaman, T. Shukui, and F. Sajjad. Exploring the root causes of terrorism in South Asia: everybody should be concerned. *Quality & Quantity*, 48(6):3065-3079, (2013).

[18] R. Pan. Rebellion on sugarscape: case studies for greed and grievance theory of civil conflicts using agent-based models. *Social Computing, Behavioral-Cultural Modeling and Prediction Lecture Notes in Computer Science*:333–340, (2011).

[19] S. Huntington. *The clash of civilizations and the remaking of world order*. New York: Simon & Schuster, (1996).

[20] P. Bernholz. International political system, supreme values and terrorism. *Public Choice*, 128: (1-2):221–231, (2006).

[21] P. Manigart. Ethnic conflict and terrorism: the origins and dynamics of civil war. *Armed Forces & Society*, 33(1):130–132, (2006).

[22] E. Bueno de Mesquita. The quality of terror. *American Journal of Political Science*, 49(3):515-530, (2014).

[23] K. Dalacoura. *Islamist Terrorism and Democracy in the Middle East*, Cambridge: Cambridge University Press, (2011).

[24] A. Samuel. Some studies in machine learning using the game of checkers. *Computer Games I*:335–365, (1988).

[25] L. Qingjie, X. Lingyu, Y. Jie, W. Lei, X. Yunlan, S. Suixiang, and L. Yang. *Research on domain knowledge graph based on the large scale online knowledge fragment.* 2014 IEEE Workshop on Advanced Research and Technology in Industry Applications (WARTIA), (2014).

[26] P. Domingos. A few useful things to know about machine learning. *Communications of the ACM*, 55(10):78, (2012).

[27] Y. Abu-Mostafa, "The Linear Model II," in *Learning from Data*. California Institute of Technology. [Online].

[28] S. Sayad. Naïve Bayesian, from *Predicting the Future.* [Online],

[29] N. Shahadat and B. Pal. An empirical analysis of attribute skewness over class imbalance on Probabilistic Neural Network and Naïve Bayes classifier. *2015 International Conference on Computer and Information Engineering (ICCIE)*, (2015).

[30] D. J. Hand and K. Yu. Idiots Bayes: Not So Stupid after All? *Revue Internationale de Statistique*, 69(3) :385, (2001).

[31] "Classification, LDA" in *Data Mining and Analysis*. Stanford University. [Online].

[32] E. Fix and J. Hodges. Discriminatory analysis: nonparametric discrimination: Consistency properties. *PsycEXTRA Dataset*, (1951).

[33] A. Padmanabha and C. Williams. "*k*-nearest neighbors." *Brilliant.org*. [Online].

[34] N. Gayar, F. Schwenker, and G. Palm. A Study of the robustness of knn classifiers trained using soft labels. *Artificial Neural Networks in Pattern Recognition Lecture Notes in Computer Science*: 67–80, (2006).

[35] S. Kaplan. A typology of terrorism. *Journal of Political Philosophy*, 6(1):1-38, (2008).

[36] "Misclassification: effects, control, and adjustment." *Statistical Methods for Rates and Proportions Wiley Series in Probability and Statistics*, p. 561–597.

[37] "Random forest - modeling the titanic voyage with R." *DataTons Blog*, 08-Jun-2017. [Online].

[38] L. Breiman, M. Last, and J. Rice. Random forests: finding quasars. *Statistical Challenges in Astronomy*, 41(1):243–254. (2001).

[39] C. Fabio and C. Ira. Risks of semi-supervised learning: how unlabeled data can degrade performance of generative classifiers. *Semi-Supervised Learning*: 56–72, (2006).

[40] F. Cucker and D. Zhou. *Learning Theory*. Cambridge: Cambridge University Press: 127–133, (2007).

[41] “Overfitting in machine learning: what it is and how to prevent it.” *EliteDataScience*, 08-Feb-2018. [Online].

[42] G. LaFree and L. Dugan. Introducing the global terrorism database. *Terrorism and Political Violence*, 19(2):181–204, (2007).

[43] B. Schneier. “Why data mining won't stop terror.” *Wired*, 04-Jun-2017. [Online].

[44] F. Ding, Q. Ge, D. Jiang, J. Fu, and M. Hao. Understanding the dynamics of terrorism events with multiple-discipline datasets and machine learning approach. *Plos One*, 12(6), (2017).

[45] C. Dwork, M. Hardt, T. Pitassi, O. Reingold, and R. Zemel. Fairness through awareness. In Innovations in Theoretical Computer Science Conference (ITCS): 214–226, (2012).