Improvements on the quality of the automatic anticipation with Pandora II

Denitsa Pesheva

ANR: 234913

Thesis submitted in partial fulfillment of the requirements for the degree of
Master of Science in Communication and Information Sciences,
Master Track Human Aspects of Information Technology,
at the Faculty of Humanities
of Tilburg University

Thesis committee:

Supervisor: Dr. Ir. P.H.M. (Pieter) Spronck

Second Reader: Dr. (Marie) Postma

Tilburg University
School of Humanities
Department of Communication and Information Sciences
Tilburg center for Cognition and Communication (TiCC)
Tilburg, The Netherlands
September, 2017

Abstract

The goal of this research is to investigate possible improvements for automatic anticipations with Pandora II database. Pandora II is a scenario-based model developed by the Dutch National Police Force (KLPD) in order to provide insights into terrorists' behavior. The results of previous studies with Pandora II have shown that data mining algorithms can be used to anticipate future events.

Since Pandora II is still subject to further development, there might be possibilities of improvements. In this study, we investigated whether or not missing values can be imputed, and which typical patterns can be identified. First, we ran preliminary experiments with ZeroR, Decision tree, KNN, SMO and Random forests algorithms on a data sample (about 13% of the whole database). Second, we used the full dataset tuned to our findings from the preliminary experiments. For Random forests and SMO, however, processing the full dataset was limited by time restrictions of the present study; therefore, we only ran ZeroR, KNN, and J48. Since the results for J48 on the extract of the dataset were mostly similar to those achieved with Random forests and SMO, we assume that they are also similar for the full dataset. Third, we ran K-means and K-modes clustering algorithms. Comparing the results of experiments to a baseline ZeroR, we found that Decision tree and SMO are more suitable for prediction of missing values in Pandora II than KNN and Random forests. The results of clustering algorithms showed that protagonists in Pandora II can be separated into two groups -- one with terrorist groups that kill more victims and one with terrorist groups that injure more victims. With regards to weapon types they use, terrorist groups were grouped as similar but separated by antagonist, area and modus operandi.

The main conclusion is that automatic anticipations with Pandora II database can be enhanced to a large extent in two ways. First, data mining techniques can be used to impute missing values, which will improve accuracies of future analyses. Second, patterns were identified in Pandora II. These findings can be used for development of new anticipation strategies.

Table of Content

Chapter 1 Introduction

In this chapter, we present the motivation for this study (1.1), followed by a problem statement and research questions (1.2). In the last section (1.3), we give a short outline of the thesis.

1.1. Motivation

Contemporary societies over the world face a constant risk of terrorist attacks. Undoubtedly, terrorism is among the gravest of threats. Both government and private sectors have allocated extensive resources to its anticipation. More developments on knowledge and insights into the phenomenon are needed in order to counter it. To date, there are a lot of studies devoted to terrorism, but scientists have not reached a consensus on its definition. According to Alex Schmid, a scholar who has studied more than one hundred explanations of the term, terrorism is a reflection of a personal political interest, and moral judgments of those who are defining it (Schmid, 2004. Malkki & Toivanen, 2010). Besides, the evolving nature of terrorism has increased the importance of finding new methods for analyzing motives, structures, and behaviors of terrorist groups (Jackson, 2005).

One relatively new approach which benefits anticipation of terrorism is a scenario-based model. Representative for this model is that collected information from previous terrorist attacks can be used as a basis for countering future violence (Carroll & Go, 2004, Gorr & Harries, 2003, Gorr, Olligschlaeger & Thompson, 2003, Khalsa, 2004). Narrative elements from movies, drama and literature scripts are used to benefit consideration of different viewpoints and definition of alternatives in decision-making processes (Carrol & Go, 2004). The protagonist, antagonist, and modus operandi, which are part of the model, can serve as useful tools for investigation of relevant constraints and changes in criminal behavior. Also, one can examine whether or not a change in one of the attributes could influence other attributes. When a relationship between different components is found, this information can be used by various institutions for terrorism anticipation (Stege, 2012).

The Pandora II is an ESC12 scenario-based model based on a database of records from previous terrorist incidents. The model was developed by The Dutch National Police Force (KLPD) in order to provide insights into terrorists' behavior. The database has already been used by Linda

Stege (2012), Sophie Bressers (2012) and Peter De Kock (2014) in their studies of terrorism anticipation with scenario-based models.

In her master thesis, Stege (2012) has investigated to what extent data mining techniques such as Decision tree, Naïve Bayes, and K-Nearest Neighbors algorithms can be used for prediction of terrorist behavior. The results have shown that the algorithms can be used to anticipate future events accurately. Bressers (2012) has explored the Pandora II model in order to highlight weaknesses and possible improvements. Research has resulted in formulation and implementation of a set of technical and conceptual improvements. She has also found that there are significant relations between *Target Type, Type of Incident, Weapon Type, Subweapon Type, and Terrorist Group* attributes, and these attributes can be considered as crucial for criminal investigations.

Since Pandora II is still subject to further development, there might be possibilities to improve the quality of automatic anticipation with the database (Stege, 2012). From previous studies, we already know that Pandora II is suited to statistical analyses, and relationships between attributes have been found. Knowing the value of one of the attributes increases chances of a correct prediction of the value of others attributes. To date, Pandora II has not been thoroughly analyzed for occurring patterns. In addition, there are a lot of missing values which may influence anticipation in general. Therefore, the primary focus of present thesis is to investigate whether or not anticipation with Pandora II can be improved. The results of this research can be used by anti-terrorism organizations to determine anticipation strategies quickly or to support human decision-making.

1.2. Problem statement and Research Questions

So far, anticipation of future criminal events with scenario-based models has shown to be successful. The assumption is that Pandora II database will be able to anticipate terrorism as well (de Kock, 2014). To continue improving the database, in this thesis we focus on the quality of automatic anticipation with Pandora II. The research objective is to investigate whether or not missing values can be predicted and imputed. Also, we investigate what typical patterns in the database can be identified. The following problem statement has been formulated:

> *“Can the quality of the automatic anticipation of terrorist activity with Pandora II be improved?”*

Two research questions are defined in order to answer the problem statement.

RQ1:

How can missing values in Pandora II be imputed?

In order to investigate RQ1, we selected twenty-three attributes, preprocessed them and ran five data mining algorithms in order to predict missing values.

RQ2:

What typical patterns occur in Pandora II?

To answer RQ2, we selected six attributes and ran clustering algorithms K-means for numeric, and K-modes for categorical type attributes.

1.3. Outline

The thesis is organized as follows: Chapter 2 presents a literature review on concepts such as crime anticipation, data mining techniques, and a definition of terrorism and scenario–based model. Chapter 3 describes data preprocessing and data mining algorithms used to answer research questions. Chapter 4 reports the results of algorithms. Finally, in Chapter 5, we answer the research questions.

Chapter 2: Background

Chapter 2 is divided into four sections. First, we explain the concept of crime anticipation (2.1). Second, we discuss data mining techniques (2.2). The third section provides a brief overview of terrorism (2.3), and finally, the fourth section presents terrorism anticipation with scenario-based models (2.4).

2.1 Crime anticipation

Crime anticipation is a relatively new concept, used by authorities to forecast criminal activities. For a long time, crime anticipation was considered infeasible, probably because its main targets were considered to be on a desired scale too small for observation, therefore not leading to a reliable model estimation (Gorr & Harries, 2003). Recently, a combination of factors such as criminological theories, development of information technologies, innovations in geographic information systems (GIS), and crime management practices have made crime anticipation relevant aspect to investigations. In their study "*Introduction to crime forecasting*", Corr and Harries (2003) pointed that the main focus on crime forecasting has changed from persons who commit crimes to places where these activities tend to occur. They explained that as a result of two reasons. First, the routine activities (Cohen & Felson, 1979), the ecology of crime (e.g., Brantingham & Brantingham, 1984), and the hot spot (spatial clusters of crime) theories (Sherman, Gartin, & Buerger, 1989) have established criminality of spaces. Second, geographic information systems (GIS) have made crimes mapping possible.

There is a great deal of uniqueness and randomness associated with crimes, but patterns can be observed as well, and they are what makes an accurate anticipation possible. In their paper "*Short-term forecasting of crime*," Gorr and Harries (2003) stated that crime is quite a seasonal occurrence. For instance, property crime levels increase late in the year (holidays are indicated as the primary prerequisite), while during cold weather, burglary and robbery crimes are more likely to increase due to seasonal economic pressures or unemployment rates. Applying crime maps and expecting that patterns will remain, police have begun forecasting one period forward. By implementation of a hot spot method (longitudinal clusters of crimes), investigators can determine crime rates of certain places (Kelling, Coles, & Wilson, 1998; Langworthy & Jefferis, 2000; Sherman, Gartin, & Buerger, 1989; Wilson & Kelling, 1982). This results in a prevention of incidents by making relevant crime alerts or targeting patrols in more dangerous areas (Gorr & Harries, 2003).

Three types of crime forecasting can be distinguished on the basis of planning horizon, and each type provides different benefits. The first one is short-term planning used for tactical deployment. Applied within small geographical areas, it benefits municipal police conducting surveillance for deployment of special units. The second type is medium-term used in resource allocation planning. And the third one is long-term planning, implemented in strategic planning and crime reduction policies. At this level, police and governments are engaged in a dialogue about priorities, or budget planning for additional forces, shift resources between prevention and enforcement activities, etc. (Gorr & Harries, 2003).

The events of September 11, 2001, have significantly increased concerns about national securities. In order to be able to prevent future attacks, authorities have been collecting domestic and foreign intelligence actively. Local forces have been motivated to monitor more closely criminal activities in their jurisdictions (Chen et al., 2004). Collected information has called for the necessity of a new generation of computational tools to support humans in extracting useful observations from the rapidly growing volume of digital data (Fayyad, Piatetsky-Shapiro, & Smyth, et al., 1996). Technology developments have made data mining techniques an easily available and powerful instrument for criminal investigators. In the next section, we will briefly discuss them.

2.2. Data mining techniques

Data mining techniques make sense of data, extracting high-level knowledge from low-level data in the context of large data sets. Being fast, powerful and user-friendly, they increase both the speed of analysis and its depth (McCue, 2005).Techniques such as classification, cluster analysis, and prediction are used for identification of patterns in structured crime data. Latest techniques are able to identify patterns from both structured and unstructured data. Recently, automated algorithms are increasingly developed and implemented in local law enforcement and national security solutions. For instance, different data bases of images or texts are analyzed by entry extraction. Clustering techniques may distinguish the various crime groups and their behavior by grouping data items into classes with similar characteristics. Classification uses predefined classes (formed on the basis of common properties among entities) to organize crime activities into groups. However, classification techniques need a predefined classification scheme, and reasonably complete training and testing data (Chau et al. 2004).

In 2005, McCue investigated how data mining techniques can be implemented for crime anticipation. He stated that by using predicting analyses, one can model complex interactions or relations accurately, and apply them for identification and characterization of unknown relationships, even to make accurate forecasts of future events. Detecting a crime with a potential risk of escalation at an early stage may prevent serious incidents to occur. In his paper "*Data Mining and Crime Analysis in the Richmond Police Department*", McCue (2010) presented examples of data mining techniques that have been implemented into practice in order to enhance decision-making and analysis at local levels. With applications, such as tactical crime analysis, risk and threat assessment, the Richmond department increased public safety significantly, as measured by a 46% reduction in citizen complaints about random gunfire and 246% increase in weapons recovered.

What makes data mining techniques a powerful instrument for criminal investigators is that they do not require extensive training skills as data analysts to be able to explore large databases quickly and efficiently. Compared to traditional tactical crime analysis they can even reduce department and personnel costs (Chau et al. 2004).

2.3 Terrorism

Terrorism is a form of criminal behavior (Chibelushi, Sharp & Shah, 2006). It overlaps with organized crime, and different transformations between them are possible. The word "Narco-terrorism" is widely used amongst authorities to explain both tactics of drug trafficking organizations and the means of funding terrorist organizations. In their paper "*Crime and terrorism*", Grabosky and Shohl (2010) illustrated the close link between terrorism and organized crime giving examples of:

- terrorist organizations that are engaged in crime to support themselves or their operations;
- terrorist organizations that may abandon their ideology and become criminal organizations;
- terrorist organizations that are teaching criminal organizations;
- terrorist organizations that are effective criminal organizations.

Loza (2007) has pointed that a terrorist act is a calculated use of unexpected, shocking, and criminal violence against non–combatants, in order to intimidate or coerce a government or civilian population to accept demands on behalf of an underlying ideology or cause. Researchers have indicated at least five goals of terrorist attacks (De Kock, 2014):

- to decrease people's confidence in their governments;
- to disorganize routine social activities;
- to create panic, chaos, fear, or paranoia;
- for revenge;
- to convey a message.

According to the Alex Schmid, a scholar in the field of terrorism research, the definition of the term is a reflection of the political interests and moral judgment of people who define it (Schmid, 2004). Together with Jongman, he proposed following definition that has been believed to provide a strong academic basis:

"Terrorism is an anxiety-inspiring method of repeated violent action, employed by (semi)clandestine individual, group or state actors, for idiosyncratic, criminal or political reasons, whereby - in contrast to assassination - the direct targets of violence are not the main targets." (Schmid & Jongman, 1988).

However, there is little consensus on how terrorism as a phenomenon should be defined. If researchers agree on anything, it is the poor quality of studies thus far (Malkki & Toivanen, 2010). An implicit assumption is that if one knows what causes terrorism, one could prevent it. And by knowing what makes an individual a terrorist, it will be easier to identify individual terrorists. In order to be able to avoid future attacks, several international organizations have created definitions that suit their needs. Van der Heide (2011) compared definitions from the US State Department, the European Union, Dutch Law and Global Terrorism Database with the Schmid and Jongman definition (1988). She concluded that all agreed on intentional use of violence, but continued debating factors such as whether victims of terrorism must be (non)combatants; or whether or not a political motive is required.

2.4 Terrorism anticipation with a scenario–based model

2.4.1 Scenario-based models

A scenario-based model is an approach that tries to predict future possibilities by both scenarios and a combination of anticipatory tools. The approach is useful in situations where constraints or changes in the external environment are recognized, but not well understood. Applied to criminal behavior, the model provides an opportunity to learn from the past and to adjust a chosen strategy. In order to be successful, it should be able to identify significant events, key players, and their motives. Scenario-based models have at least two advantages. The first one is that multiple cases within a model can be compared for similarities and differences. The second advantage lies in a knowledge management notion: the storage of knowledge and experience. Organizations need to store tacit and explicit knowledge to prevent a loss of significant knowledge and experience when experienced employees retire or resign (Debowksi, 2006).

2.4.2 Terrorism anticipation with Pandora II

Terrorist attacks have been described as scenarios or as a collection of story elements. Scenario–based models identify behavioral patterns within criminal organizations and increase opportunities for anticipating criminal behavior (Bressers, 2012).

Pandora II is a database that combines a conceptual design of the ESC12 scenario-based model (see Table 1) with a dataset of terrorist incidents between 1970 and 2015. It was created “to enable comparison of large amounts of data and to provide insight into processes of radicalization and terrorist planning” (Van der Heide, 2011). The ESC12 is a set of twelve Elementary Scenario Components that are building blocks of each imaginable scenario. They are characterized by meaning and by relation, and every individual component has a dynamic relation with other components (de Kock, 2014).

Table1 *A brief description of the twelve Elementary Scenario Components (ESC12)*

#	Component	Description
1.	**Protagonist**	The perpetrator of an incident
2.	**Antagonist**	The victim of an incident (which is not necessarily a person but can also be a building or object)
3.	**Context**	The circumstances under which the incident took place.
4.	**Arena**	The location where the incident took place.
5.	**Time**	The moment at which the incident took place.
6.	**Modus Operand**	The actions that happened before, during and after the incident.
7.	**Means**	The sources that have been used for the incident
8.	**Motivation**	The reason why the protagonist has offended the antagonist.
9.	**Primary Purpose**	Where the protagonist was striving for with the incident
10.	**Resistance**	The obstacles the protagonist had to overcome to be able to perform his act
11.	**Red Herring**	A misleading occurrence or indicator often used to lead someone in the wrong direction or to make someone believe an untruth.
12.	**Symbolism**	Occurs when a specific act is of symbolic value for the offender, the victim, an audience or another specific individual or group of Persons

Chapter 3: Methodology

In the first section of this chapter, we explain data preprocessing (3.1). In the second section (3.2), we present the five classifiers that we have used for prediction of missing values. In the last section (3.3), we discuss clustering algorithms that we have run.

3.1 Data preprocessing

3.1.1 Attributes selection

To prepare the data for data mining and analyses, it usually needs to undergo preprocessing. In Pandora II, the database that we used, there are about 150,000 records with 134 attributes, coded as numeric, categorical and text types. In order to select attributes that can be utilized, we set two criteria.

The first criterion was an attribute's type. For data mining algorithms that will be used to identify patterns and to fill missing values, attributes can be numeric and categorical types. The third category, text type attributes, will not be considered as they contain highly diverse records that cannot be transformed into numeric or categorical. This eliminated 62 of the attributes.

The second criterion was the total number of records in which a particular attribute is not missing, because as Gorr and Harries (2003) pointed, smaller samples increase forecast errors. Attributes with 10,000 or less records might not be able to provide reliable results, so they were excluded. The number of attributes that will be used for our experiments was reduced to 23 (141,966 records). Table 2 presents information about attributes – their types and the number of categories that they have. For a full description of the attributes, see Appendix A.

Table 2 *A list of selected attributes.*

#	Attribute	Type of the attribute	Number of categories
1	Year	Numeric	44
2	Month	Numeric	12
3	Day	Numeric	31
4	Country	Categorical	203
5	Region	Categorical	12
6	Extended incident	Categorical	2
7	Vicinity	Categorical	2
8	Inclusion Criteria 1	Categorical	2
9	Inclusion Criteria 2	Categorical	2
10	Inclusion Criteria 3	Categorical	2
11	Part of Multiple Incident	Categorical	2
12	Successful Attack	Categorical	2
13	Suicide Attack	Categorical	2
14	Attack Type	Categorical	8
15	Target/Victim Type	Categorical	21
16	Target/Victim Subtype	Categorical	111
17	Nationality of Target/Victim	Categorical	207
18	First Perpetrator Group Suspected/Unconfirmed	Categorical	2
19	Weapon Type	Categorical	11
20	Total Number of Fatalities	Numeric	156
21	Total Number of Injured	Numeric	217
22	Hostages or Kidnapping Victims	Categorical	2
23	International- Miscellaneous	Categorical	2

3.1.2 Preprocessing of the attributes

To preprocess selected attributes, we used R-Studio Desktop 1.0.143. The numeric attributes were transformed into nominal because most of the classifiers can handle or perform better with categorical types only. After that, attributes were discretized. Discretization is a form of data transformation, where raw attributes' values are replaced by ranges, which can impact classifiers' performance on high-dimensional data significantly (Witten et al., 2011).

The following categorical attributes had a level called "Unknown" that has been removed because it does not provide any useful information: *Target/Victim Type*, *Weapon Type*, *Attack Type*, *Vicinity*, *Hostages or Kidnapping Victims*, and *International - Miscellaneous.* All rows with at least one missing value have been removed. The total number of records became 112,594.

3.1.3. Preliminary experiments for Research Question 1

Classifiers such as K-Nearest Neighbors (KNN), Sequential Minimal Optimization (SMO), and Random forests take time to be applied to large samples. As this study is limited by time, we decided to run preliminary experiments with a subset of the data. We made a new sample of 14,197 records or about 13% of the whole data, resampling it in a random way. After the pretests, we ran the Decision tree and KNN classifiers again, because they proved to be effective, but this time, we used the whole data set. Table 3 contains a list of target attributes with missing values that were classified. The remaining 11 attributes had no missing values. Each attribute was classified on the basis of all other attributes. The classifiers were applied to the training set (80% of the records), and test set (20% of the records).

Table 3 *Target attributes*

#	Attribute	Number of missing values	Type of the attribute	Number of categories
1	Month	23	Numeric	12
2	Day	896	Numeric	31
3	Vicinity	67	Categorical	2
4	Attack Type	4714	Categorical	8
5	Target/Victim Subtype	7114	Categorical	111
6	Nationality of Target/Victim	1035	Categorical	207
7	First Perpetrator Group Suspected/Unconfirmed	378	Categorical	2
8	Weapon Type	11046	Categorical	11
9	Total Number of Fatalities	8172	Numeric	156
10	Total Number of Injured	12758	Numeric	217
11	Hostages or Kidnapping Victims	331	Categorical	2
12	International- Miscellaneous	486	Categorical	2

3.2 Classifiers

In order to answer RQ1 and to propose a best-predicting classifier to impute missing values in Pandora II, we tested five well-known classifiers: ZeroR, Decision tree J48, K-Nearest Neighbors (KNN), Sequential Minimal Optimization (SMO), and Random forests. The classifiers were run in a data mining tool Weka, version 3.6.15. In the following subsections, we explain the four classifiers briefly.

3.2.1 Decision tree

Decision tree is a flowchart-like structure where each internal node designates a test on a particular attribute. The test outcome is represented by a branch, and a leaf node grasps a class label. As a root node, a tree takes the first node. Decision trees could be binary (every branch has exactly two nodes) or non-binary (more than two nodes per branch). In order to select an optimal split and to remove sections that contribute less to the classification of instances, a decision tree algorithm uses entropy reduction. It attempts to identify and remove branches that might reproduce noise or outliers in the training data. This algorithm is used a lot in data mining because it does not require any domain knowledge, can handle multidimensional data and, in general, has good accuracy (Witten et al., 2011).

3.2.2 K–Nearest Neighbors (KNN)

The K-Nearest Neighbors (KNN) algorithm is an instance-based classification algorithm in which a new instance is compared (on the basis of the distance) to the most similar existing training examples in an n-dimensional space. More than one nearest neighbor is used, and the majority class of the closest k neighbors (or the distance weighted average if the class is numeric) is assigned to the new instance (Witten et al., 2011).

3.2.3 Sequential Minimal Optimization (SMO)

Sequential Minimal Optimization is a supervised algorithm used for training of Support Vector Machines (SVM). In order to solve the SVM's quadratic programming (QR) optimization problem, it decomposes a large QR problem into a set of small QR problems and solves them analytically. This makes the SMO faster and allows it to handle large and sparse datasets (Witten et al., 2011).

3.2.4 Random forests

Random forests are an ensemble algorithm used for classification or regression tasks. It generates a lot of decision trees using randomly selected attributes at each node in order to determine the split. The output of classification is that class which is the mode of the classes by all individual decision trees. The algorithm is indifferent to the number of selected attributes and overfitting (Witten et al., 2011).

3.2.5 Classifiers Setup

The ZeroR classifier was used as a baseline to determine the benchmark for the other classifiers. This classifier merely predicts a majority category (class). The classifiers were run on all 23 attributes.In KNN, we adjusted the number of nearest neighbors to 3 because the pretest showed k=3 as the most effective. The criterion for measurement of classifiers' success levels, we used the performance accuracy, also called confidence. It refers to the percentages of correctly classified instances of all cases to which the classifier applies (Witten et al., 2011).

3.3 Clustering algorithms for Research Question 2

Clustering is an unsupervised algorithm used for grouping data observations into groups (clusters) based on how similar they are. The observations in a particular cluster are similar to each other, but dissimilar to observations in other groups (Han, Kamber, & Pei, 2012).

To investigate RQ2, first, we created a data subset with following attributes that have the potential to reveal information about protagonists and their modus operandi: *Total Number of Fatalities*, *Total Number of Injured*, *Attack Type*, *Target/Victim Subtype*, *Weapon Type*, and *Country*. Then, we selected the eleven most frequently occurring groups of all 3189 terrorist groups in Pandora II (see Table 4). The first column gives information on a terrorist group's name. The second column shows Frequency of group's occurrence, while the third presents percentages with regard to the entire database. The data subset had 19,995 records.

Table 4 *The top eleven the most occurring groups in Pandora II*

#	Group Name	Frequency	Percentages
1	Unknown	54912	48.77%
2	Shining Path (SL)	3626	3.22%
3	Taliban	3421	3.04%
4	Farabundo Marti National Liberation Front (FMLN)	2438	2.17%
5	Revolutionary Armed Forces of Colombia (FARC)	1757	1.56%
6	Basque Fatherland and Freedom (ETA)	1673	1.49%
7	New People's Army (NPA)	1548	1.37%
8	Irish Republican Army (IRA)	1544	1.37%
9	Communist Party of India - Maoist (CPI-Maoist)	1334	1.18%

10	Liberation Tigers of Tamil Eelam (LTTE)	1333	1.18%
11	Islamic State of Iraq and the Levant (ISIL)	1251	1.11%

We noticed that the biggest group was ‘Unknown’, which holds all cases for which the name of the terrorist group was not available, and removed this category. To identify patterns, we ran K-means clustering for numeric features, and K-modes for categorical ones. Table 5 presents selected attributes for K-means, while Table 6 lists these for K-modes.

Table 5 *Selected attributes for K-means*

#	Attribute	Type of attribute
1	Total Number of Fatalities	Numeric
2	Total Number of Injured	Numeric

Table 6 *Selected attributes for K-modes*

#	Attribute	Type of attribute
1	Attack Type	Categorical
2	Target/Victim Subtype	Categorical
3	Weapon Type	Categorical
4	Country	Categorical

In the following subsections, we describe the clustering algorithms that we used.

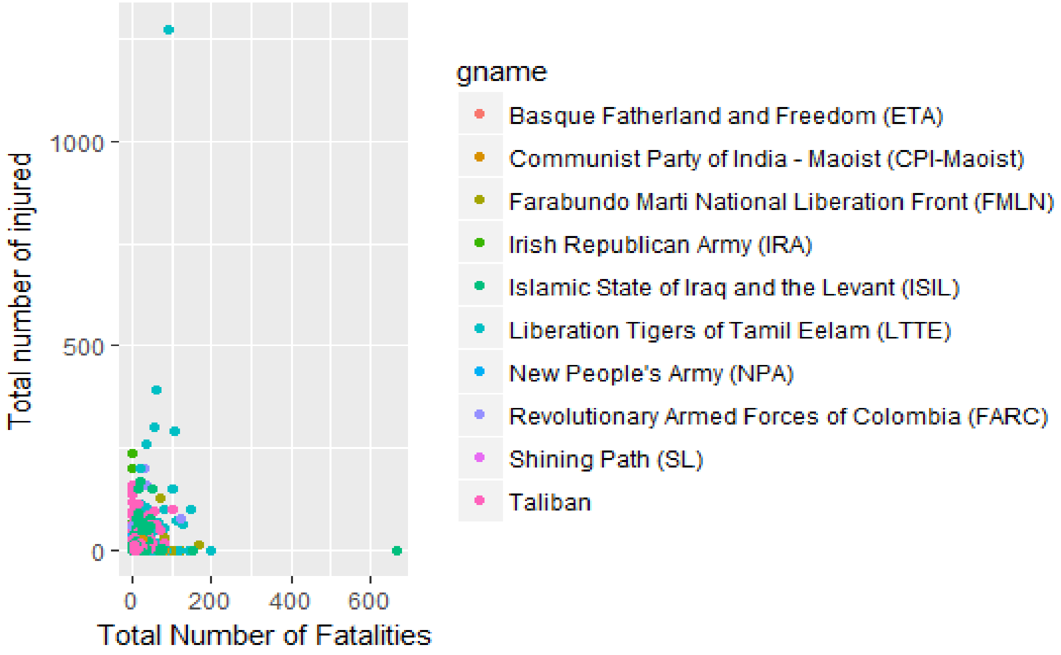
3.3.1 K-means clustering and K-modes

K-means is a centroid-based technique, where a centroid of a cluster is its center point, defined as a mean value of the observations within it. To measure similarities, the algorithm uses the Euclidean distance between observations and the cluster mean. Because of that, K-means is suitable for numeric attributes, when the mean of a set of objects is defined. A disadvantage of K-means is the necessity for users to determine the number of clusters beforehand. A common practice to overcome it is to apply the Elbow method. The Elbow method takes the percentage of variance explained as a function of the number of clusters. The assumption is that if the number of clusters is increased, that can help reducing of the sum of within-cluster variance of each cluster (Han et al., 2012). K-modes clustering is an extended variant of K-means for clustering of categorical (nominal) attributes. It takes modes of clusters instead of their means (Han et al., 2012).

3.3.2 Clustering Setup

During the Exploratory Data Analysis, we noticed that there were outliers in *Total Number of Fatalities* and *Total Number of injured attributes* (see Figure 1).

Figure 1 *Exploratory Data Analysis of attributes*



The outliers could bias the results of K-means (as it is based on means), so we removed these six cases - one of Islamic State of Iraq and the Levant (ISIL) terrorist group (more than 240 in *Total Number of Fatalities*), and five of Liberation Tigers of Tamil Eelam (LTTE) terrorist group (more than 240 in *Total Number of injured*). To determine the right number of clusters, we applied the Elbow method. Figure 2 shows the results for K-means, and Figure 3 for K-modes.

Figure 2 *The elbow method for K-means.*

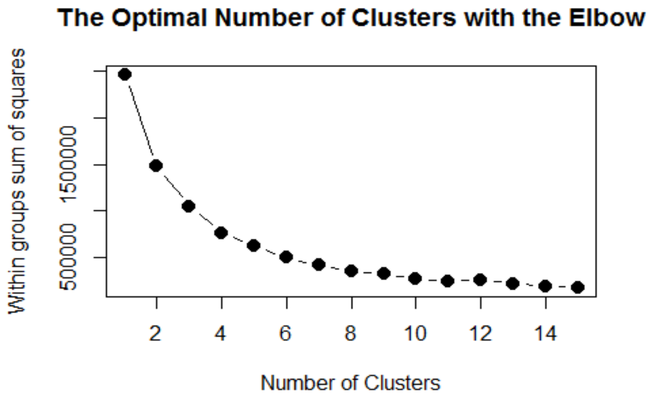
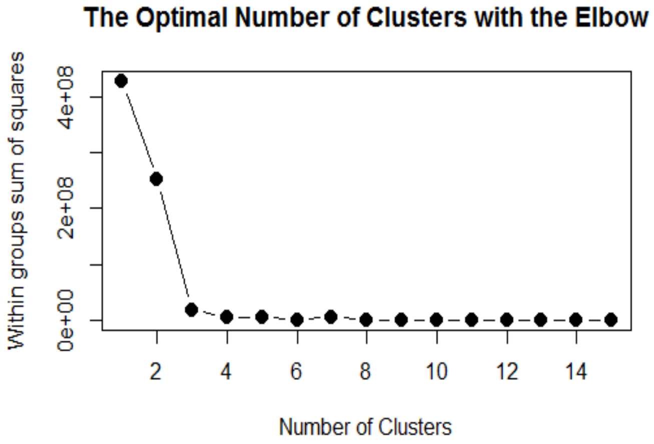


Figure 3 *The elbow method for K-modes.*



Based on these figures, we decided to use 4 as number of clusters for both K-means and K-modes (k=4).

Chapter 4: Results

In this chapter, we present the results of the classifiers and clustering algorithms. In the first section (4.1), we explain outcomes for RQ1, while in the second section (4.2) we answer RQ2.

4.1 Results Research Question 1

In order to investigate RQ1 "*How can missing values in Pandora II be imputed?*", we ran the four classifiers (described in chapter 3), and the baseline ZeroR with each target attribute. Because of time restrictions (a large dataset takes a long time to run some of the algorithms), we did preliminary experiments on a small extract: 14,197 records, or 13%, of the whole dataset. In Table 7, we present results of the tests.

Table 7 *Accuracy (in percentages) of classifiers ran on a subset of the whole dataset*

#	Target Attribute	ZeroR accuracy	J48 accuracy	KNN accuracy	SMO accuracy	Random forests accuracy
1	Month	8.99%	15.67%	13.24%	11.91%	17.57%
2	Day	3.95%	7.43%	6.23%	4.97%	8.67%
3	Vicinity	91.76%	91.76%	91.79%	92.07%	92.15%
4	Attack Type	51.07%	85.59%	81.22%	88.45%	83.48%
5	Target/Victim Subtype	6.97%	51.78%	37.87%	50.30%	46.53%
6	Nationality of Target/Victim	12.93%	92.99%	80.20%	93.38%	93.24%
7	First Perpetrator Group Suspected/Unconfirmed	89.96%	89.96%	89.99%	89.82%	90.60%
8	Weapon Type	53.26%	91.51%	87.92%	92.07%	90.74%
9	Total Number of Fatalities	55.58%	61.61%	59.39%	61.32%	58.05%
10	Total Number of Injured	62.03%	62.03%	61.57%	62.66%	62.56%
11	Hostages or Kidnapping Victims	95.91%	99.37%	97.35%	98.84%	98.10%
12	International- Miscellaneous	88.24%	97.71%	95.91%	98.06%	96.59%

The results have shown that Decision tree J48, SMO, and Random forests performed more or less the same, while KNN almost always performed worse than the other ones. We compared accuracies of the classifiers with baseline ZeroR. For the attributes *Vicinity*, *First Perpetrator Group Suspected/Unconfirmed*, and *Total Number of Injured*, the percentages were the same or slightly higher than the baseline. For *Month*, *Day*, *Attack Type*, *Target/Victim Subtype*, *Nationality of Target/Victim, Weapon Type, Total Number of Fatalities, Hostages or Kidnapping Victims,* and *International- Miscellaneous*, all four classifiers achieved better results than ZeroR. That means they predict better than the majority class.

The Decision tree J48 provides information about connections between different attributes, so the features that it uses for prediction of the target attributes can be inspected. In the following subsections, we explain Decision trees with accuracies higher than ZeroR.

4.1.1. Attack Type

The baseline classifier ZeroR classified 51% of instances correctly as Bombing/Explosions. If we examine the Decision tree of *Attack type*, we see that it first takes *Hostages or Kidnapping* as a root node. After that, the tree checks *Weapon Types*, and on the next leafs, it looks at *Target/Victim Subtype* and *Total Number of Fatalities*. The biggest class was hijacking (n=1418), followed by armed assaults (n=565). In practice, this information can be used for prediction of missing values (if there are any), and check for a pattern.

4.1.2 Target/Victim Subtype

ZeroR for *Target/Victim Subtype* classified only 198 of records correctly, and its accuracy was 6.97%. The Decision tree (51.78%) performed much better than that. The attributes that the algorithm checks are *Target/Victim* Type as a root (which is obvious) and *Nationality of Target/Victim* as a next node. Depending on the results of *Nationality of Target/Victim*, the following leafs are different. For instance, if the nationality of a target is Afghan, the next node is *Total Number of Fatalities*. If the nationality of victims is Argentine, then the tree takes the attribute *Year*. For Colombian, the Decision tree checks *Weapon type*.

4.1.3 Nationality of Target/Victim

The best performing classifier for *Nationality of Target/Victim* was Decision tree J48 with 92.99% accuracy, compared to 12.93% accuracy for ZeroR. As a first attribute, it checks *Region*, followed by *Country* as a second one. The third attribute that it takes is *International-Miscellaneous*. The Region is classified as North America (Canada, Mexico, United States), and the country is Canada. If a perpetrator group attacked a target of a different nationality, then nationality of victims is classified as Mexicans. If not, then the nationality of targets is likely to be Americans. The results suggest that targets are often Mexican or American. One possible explanation might be that the dataset was mostly built in the United States, so it may contain mostly information on attacks against US citizens. The next node is *Target/Victim Subtype*.

4.1.4 Weapon Type

The accuracy of the ZeroR baseline for *Weapon Type* is 53.26%. The Decision tree J48 achieved a much better result, namely 91.51%. As a first node, it takes *Attack Type*. The type of attack is classified as Assassination, and tree checks *Successful Attack*. Then the tree continues with *Country*.

4.1.5 Total Number of Fatalities

For *Total Number of Fatalities*, all of the classifiers achieved better results than the baseline ZeroR (55.58%). The accuracy of Decision tree J48 was 61.61%. If we check the Decision tree for *Total Number of Fatalities*, we can see that it first looks at *Weapon type*, then the second node is *Attack type*, and the third one is *Successful Attack*. In practice, if one knows *Weapon Type*, can quickly estimate the approximate number of expected victims.

4.1.6 Hostages or Kidnapping Victims

The ZeroR classified 2723 of instances as attacks without hostages, and its accuracy was 95.91%. In the root node, the Decision tree looks at *Extended Incident*.

4.1.7 International-Miscellaneous

The best performing classifier for this attribute was SMO (98.06%), followed by Decision tree with a small difference in accuracy (97.71%). The classifiers performed better than the baseline (88.24%). The first attribute that J48 looks at is *Nationality of Target/Victim*, the second node is *Country*, and then *Weapon type*.

4.1.8 Full Dataset

So far, we presented the results of classifiers that have been run on a subset of the full dataset. Naturally, results get more interesting when the full dataset is used. For Random Forests and SMO, however, processing the full dataset would take too much time, so we only ran it for ZeroR, KNN, and J48. Since the results for J48 on the extract of the dataset were mostly similar to those achieved with Random forests and SMO, we assume that they are also similar for the full dataset. Table 8 presents the new results.

Table 8 *Accuracy (in percentages) of classifiers for the full dataset*

#	Attribute	ZeroR accuracy	J48 accuracy	KNN accuracy	SMO accuracy*	Random Forests accuracy*
1	Month	9.18%	20.47%	17.30%	11.91%	17.57%
2	Day	3.33%	14.07%	9.57%	4.97%	8.67%
3	Vicinity	92.51%	92.51%	92.43%	92.07%	92.15%
4	Attack Type	51.86%	86.81%	84.00%	88.45%	83.48%
5	Target/Victim Subtype	7.06%	51.64%	42.25%	50.30%	46.53%
6	Nationality of Target/Victim	12.48%	94.11%	80.44%	93.38%	93.24%
7	First Perpetrator Group Suspected/Unconfirmed	90.30%	90.30%	90.31%	89.82%	90.60%
8	Weapon Type	54.18%	92.02%	89.44%	92.07%	90.74%
9	Total Number of Fatalities	54.12%	63.08%	62.46%	61.32%	58.05%
10	Total Number of Injured	62.70%	62.70%	62.67%	62.66%	62.56%
11	Hostages or Kidnapping Victims	96.00%	99.16%	97.82%	98.84%	98.10%
12	International- Miscellaneous	88.97%	97.97%	95.66%	98.06%	96.59%

*Presented results of SMO and Random Forests are on the basis of the 13% of the full dataset.

The results are comparable with those achieved on the extract, and indicated the classifiers that have been developed in this study can be used to impute missing values for *Attacktype, Target/Victim Subtype, Nationality of Target/Victim, Weapon Type, Total Number of Fatalities, International- Miscellaneous* attributes. The user needs to take into account that the imputed values are not certain – the accuracy may be considered the certainty, though a closer examination of a decision tree may indicate that some imputed values may be more certain than that (for instance, in a tree branch where there are no wrong classifications, certainty is probably very high).

4.2 Results Research Question 2

To answer Research Question 2: '*What typical patterns occur in Pandora II?*' we selected *Total number of Fatalities, Total Number of Injured, Attack type, Target type, Weapon type,* and *Country* attributes, for the ten most occurring protagonists. The knowledge of modus operandi of a terrorist group can help anti-terrorist organizations for developing strategies for anticipation. We ran two clustering algorithms to investigate what protagonists have in common, and what distinguishes them. In subsection 4.2.1 we will present results of K-means clustering. In subsection 4.2.2 we will present the results of K-modes clustering.

4.2.1 Results K-means clustering

We ran K-means with two numerical attributes – *Total number of Fatalities,* and *Total Number of Injured*, with 4 clusters (k=4), because the Elbow method (described in section 3.3.1) showed 4 clusters as the optimal number for our dataset. Table 9 reports the results of K-means clustering. The first column gives information about cluster's number. Columns two and three show the centers of clusters, while column four presents sizes of clusters. In columns five to fourteen are listed names of terrorist groups.

Table 9 *Results K-means clustering*

Cluster	Total number of Fatalities	Total Number of Injured	Cluster size	SL	Taliban	FMLN	FARC	ETA	NPA	IRA	CPI	LTTE	ISIL
1	**33.85**	2.70	402	125	37	95	17	-	18	1	9	83	17
2	6.81	**20.48**	1043	61	254	87	59	41	43	58	18	137	285
3	1.53	0.99	**18364**	3439	3102	2252	1672	1630	1485	1477	1306	1074	927
4	**26.96**	**89.61**	109	1	28	4	9	2	2	8	1	33	21
Total number of cases per group				3626	3421	2438	1757	1673	1673	1544	1334	1327	1250

SL= Shining Path**, FMLN**= Farabundo Marti National Liberation Front, **FARC**= Revolutionary Armed Forces of Colombia, **ETA**= Basque Fatherland and Freedom, **NPA**= New People's Army, **IRA**= Irish Republican Army, **CPI**=Communist Party of India – Maoist, **LTTE**=Liberation Tigers of Tamil Eelam, **ISIL**= Islamic State of Iraq and the Levant

The biggest cluster was cluster 3 with 18364 observations, and that one indicates 0-1 fatalities and 0-1 injured. Moreover, clusters 2 and 4 showed more injured than fatalities after a terrorist act, while only clusters 1 and 3 showed more fatalities than injured. The results indicated that whether more victims were killed or injured, the protagonists can be separated into two groups – those that killed more victims, and those that injured more victims.

4.2.2 Results K-modes

For K-modes, we selected four categorical attributes – *Attack type*, *Target/Victim type*, *Weapon type,* and *Country*, and ran the algorithm with four clusters (k=4). Table 10 provides information about the results of K-modes clustering. The first column shows cluster's number. The columns from 2 to 5 list modes of clusters, while column 6 presents their sizes. The columns from 7 to 14 give information about cases per terrorist group.

Table 10 *Results of K-modes clustering*

Cluster	**Attack type***	**Target/Victim type***	**Weapon type***	**Country***	**Cluster size**	SL**	**Taliban****	**FMLN****	**FARC****	**ETA****	**NPA****	**IRA****	**CPI****	**LTTE****	**ISIL****
1	1	3	5	159	5071	1735	761	84	391	520	543	392	296	238	111
2	3	14	6	4	9836	1569	2180	1059	785	1031	256	804	521	575	1056
3	7	1	8	603	1089	72	87	67	124	76	181	245	8	222	7
4	2	4	5	61	3929	250	393	1228	457	46	568	103	295	512	77
Total number of cases per group						3626	3421	2438	1757	1673	1548	1544	1120	1547	1251

***Attack type** (1= Assassination, 3= Bombing/explosion, 7= Facility/ infrastructure attack, 2= Armed assault), **Target/Victim type** (3= Police, 14= Private citizens & property, 1= Business, 4= Military), **Weapon type** (5= Firearms, 6= Explosives/Bombs/Dynamite, 8= Incendiary), **Country** (159= Peru, 4= Afghanistan, 603= United Kingdom, 61= El Salvador)

****SL**= Shining Path, **FMLN**= Farabundo Marti National Liberation Front, **FARC**= Revolutionary Armed Forces of Colombia, **ETA**= Basque Fatherland and Freedom, **NPA**= New People's Army, **IRA**= Irish Republican Army, **CPI**=Communist Party of India – Maoist, **LTTE**=Liberation Tigers of Tamil Eelam, **ISIL**= Islamic State of Iraq and the Levant

The biggest cluster was cluster 2 with 9836 observations, and it indicates protagonists that use explosives to attack citizens in Afghanistan. In cluster 1 are groups from Peru of which the target type is police (police offices, buildings or employees), and which also use explosives for their attacks. Terrorist groups in cluster 3 are from the United Kingdom and attack businesses mainly, with weapon type explosives as well. In the last cluster 4 are groups from El Salvador that attack Military.

The results indicated that, with regards to a weapon type (Means component of ESC12, see Table 1), protagonists are similar. The components that distinguish them are antagonist, area, and modus operandi components (see Table 1).

Chapter 5 Conclusion and Discussion

Chapter 5 provides an answer to the problem statement of this study:*"Can the quality of the automatic anticipation of terrorist activity with Pandora II be improved?"* .Research questions 1 and 2 are discussed in sections 5.1. and 5.2. A general discussion follows in section 5.3. In section 5.4 are listed limitations of the study.

5.1 Answer to Research Question 1

The first research question investigated "*How can missing values in Pandora II be imputed*?" To answer it, we ran five of the most widely used classifiers – ZeroR, Decision tree, K-Nearest Neighbors (KNN), Sequential Minimal Optimization (SMO) and Random Forests on 14,197 records of 12 of selected attributes. It should be noted that for *Vicinity, First Perpetrator Group Suspected/Unconfirmed, Total Number of Injured* attributes the classifiers did not perform better than the baseline ZeroR. However, the accuracies of the classifiers for *Month, Day, Attack Type, Target/Victim Subtype, Nationality of Target/Victim, Weapon Type, Total Number of Fatalities, Hostages or Kidnapping Victims, International- Miscellaneous* attributes were higher than the accuracy of ZeroR. Additionally, Decision tree reveals information about relationships between attributes that might be useful for terrorism anticipation in general. If investigators know which attributes are predictive for particular target attributes, they can use them as a basis of their speculations.

The current findings support the view that data mining techniques can be used for imputation of missing values for *Attacktype, Target/Victim Subtype, Nationality of Target/Victim, Weapon Type, Total Number of Fatalities, International- Miscellaneous* attributes of Pandora II database. Classifiers such as Decision tree, SMO, and Random forests are successful in classification of selected attributes. However, KNN almost always performed worse than other ones, which means that it is less suitable.

5.2 Answer to Research Question2

Research Question 2 examined "*What typical patterns occur in Pandora II?"* For that purpose, we created a subset of six attributes - *Total number of Fatalities, Total Number of Injured, Attack type, Target/Victim type, Weapon type, and Country,* for the ten most occurring terrorist groups in Pandora II. Expectations are that protagonists have their own, recognizable modus operandi; hence, there might be observed patterns. To investigate what distinguishes different terrorist groups, we ran two clustering algorithms – K-means for numerical, and K-modes for categorical attributes.

Analyses from K-means revealed that two types of protagonist can be distinguished -- one with more fatalities, and one with more injured victims after the attack (see Table 9). A possible explanation could be different weapon types that terrorist organizations use, the type of people they have recruited for the attacks, or the goals that they aim for. The result from K-modes indicated that protagonists are similar with regards to weapon types but different with regards to antagonists, modus operandi, and areas (see Table 10).

Taken together, patterns between protagonists in Pandora II occur. Terrorist groups use similar weapons, but they differ in carrying out attacks, targets that they choose, and the number of fatalities after attacks. There are organizations attacking mainly citizens (e.g. Taliban), while others are focused on businesses and infrastructures (e.g., IRA).

5.3 General Discussion

The purpose of this study was to investigate *'Can the quality of the automatic anticipation of terrorist activity with Pandora II be improved?*. Overall, the results imply that predictions with Pandora II can be enhanced to a large extent in two ways. First, data mining techniques can be used to predict missing values in *Attack Type, Target/Victim Subtype, Nationality of Target/Victim, Weapon Type, Total Number of Fatalities, Hostages or Kidnapping Victims, International- Miscellaneous* attributes. The algorithms that classified attributes effectively were: Decision tree, SMO, and Random Forests. Second, there are patterns that occur in Pandora II, and that knowledge can enhance anticipation with the database by a development of better strategies, that is expected to result in better forecasting.

5.4 Limitations

Three limitations of this study should be acknowledged. The first limitation was that the text attributes in Pandora II cannot be coded as numerical. Hence, they cannot be analyzed with statistical tests for relationships, while text attributes may still be useful for terrorism anticipation. The second limitation is that most of the attributes in the database are measured on a nominal scale, and that restricts analyses that can be applied. The third limitation is related to speed of the algorithms. For Random forests and SMO algorithms, we presented the results of classifiers that have been run on a subset of the full dataset.

Future investigations could disclose whether or not other data mining techniques, such as neural networks, can perform better on Pandora II. Moreover, anticipation could benefit from focusing on text attributes that have not been analyzed a lot to date.

References

1. Brantingham, P. J., & Brantingham, P. L. (1984). Patterns in crime. New York: Macmillan.
2. Bressers, S. (2012). Improvements to a Scenario Model for Investigation of Terrorist behavior. (Master thesis, HAIT). *Tilburg University*, Tilburg.
3. Carroll, J.M., & Go, K. (2004). The Blind Man and the Elephant: Views of Scenario-Based System Design. *Interactions. November + December*, 44.
4. Cohen, L. E., & Felson, M. (1979). Social change and crime rate trends: A routine activity approach. *American Sociological Review*, 44, 588–607.
5. Chen, H., Chung, W., Xu, J. J., Wang, G., Qin, Y., & Chau, M. (2004). Crime data mining: a general framework and some examples. *Computer*, 37(4), 50-56.
6. Chibelushi, C., Sharp, B., & Shah, H. (2006). ASKARI: A crime text mining approach. (P. Kanellis, E. Kiountouzis, N. Kolokotronis, & D. Martakos, Eds.) *Digital Crime and Forensic Science in Cyberspace*, 155 - 174.
7. Debowski, S.(2006). Knowledge management. *Milton Qld: John Wiley & Sons Australia, LTD.*
8. De Kock, P. A. M. G. (2014). Anticipating criminal behavior: Using the narrative in crime-related data Tilburg: *Tilburg center for Cognition and Communication* (TiCC).
9. Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). From Data Mining to Knowledge Discovery in Databases. *AI Magazine*, vol. 17(3), 37-54.
10. Grabosky, P., & Stohl, M. (2010). Crime and Terrorism. London: *Sage Publications Ltd.*
11. Gorr, W., Harries, R., (2003). Introduction to crime forecasting. *International Journal of Forecasting, vol.19*, 551-555.
12. Gorr, W., Olligschlaeger, A., & Thompson, Y. (2003). Short-term forecasting of crime. *International Journal of Forecasting, vol.19*, 579-594.
13. Han, J., Kamber, M., Pei, J., (2012). Data mining: Concepts and techniques, third edition (3rd ed.). Waltham, Mass.: Morgan Kaufmann Publishers. .
14. Jackson, B. (2005). Aptitude for Destruction: Organizational Learning and Terrorist Groups and its Implications for Combating Terrorism. In M. Leann Brown, Michael

Kenney and Michael Zarkin (red.) *Organizational Learning in the Global Context*. Aldershot: Ashgate

15. Kelling, G. L., Coles, C. M., & Wilson, J. Q. (1998). Fixing broken windows: Restoring order and reducing crime in our communities. *Touchstone Books, Carmichael, CA.*
16. Khalsa, S.K. (2004). Forecasting Terrorism: Indicators and Proven Analytic Techniques. *The Scarecrow Press, Inc*. 2004, 1-103.
17. Langworthy, R. H., & Jefferis, E. S. (2000). Utility of standard deviation ellipses for evaluating hot spots. In Goldsmith, V., McGuire, P. G., Mollenkopf, J. H., & Ross, T. A. (Eds.), *Analyzing crime patterns: Frontiers of practice*. Thousand Oaks: Sage Publications, pp. 87–104.
18. Loza, W. (2007). Psychology of Extremism and Terrorism: A Middle-Eastern Perspective. *Aggression and Violent Behavior,* Vol.12, Issue:2, 141 – 155.
19. Malkki, L., & Toivanen, R. (2010). Editors' introduction: terrorism – myths, agendas, and research.*Critical Studies on Terrorism,* 3:2, 243-246.
20. McCue, C. (2005). Data Mining and Predictive Analytics: Battlespace Awareness for the War on Terrorism. *Defense Intelligence Journal*, 13(1&2), 47-63.
21. McCue, C. (2010). Data Mining and Crime Analysis in the Richmond Police Department.
22. Stege, L., (2012). Anticipating terrorism with Pandora II. (Master thesis, HAIT). *Tilburg University*, Tilburg.
23. Schmid, A. (2004). Terrorism: The definitional problem. *Case Western Reserve journal of International Law*, vol.2 - 3 (36), 384.
24. Schmid, A. & Jongman, A. (1988). Political terrorism: a new guide to actors, authors, concepts, data bases, theories and literature. *Amsterdam: Transaction Books*.
25. Sherman, L. W., Gartin, P. R., & Buerger, M. E. (1989). Hot spots of predatory crime: Routine activities and the criminology of place. *Criminology,* 27, 27–55.
26. Van Der Heide, L. (2011). Individual Terrorism: Indicators of Lone Operators. Master, *University of Utrecht,* Utrecht.
27. Wilson, J. Q., & Kelling, G. L. (1982). Broken windows: The police and neighborhood safety. *Atlantic Monthly*, 249, 29–38.

28. Witten, I., Frank, E., Hall, M., (2011). *Data Mining: Practical Machine Learning Tools and Techniques.*

APPENDIX A: Description of the attributes

The following descriptions are adopted from Global Terrorism Database (GTD_CODEBOOK_2015), maintained by the National Consortium for the Study of Terrorism and Responses to Terrorism (START)

1. Day

(day)

Numeric Variable

Day contains the numeric day of the month on which the incident occurred.

2. Month

(month)

Numeric Variable

Month contains the number of the month in which the incident occurred.

3. Year

(year)

Numeric Variable

Year contains the year in which the incident occurred.

4. Country

(country)

Categorical Variable

This field identifies the country or location where the incident occurred. In the case where the country in which an incident occurred cannot be identified, it is coded as “Unknown.”

Country (Location) Codes

(Note: These codes are also used for the target/victim nationality fields. Entries marked with an asterisk (*) only appear as target/victim descriptors in the GTD.

4 = Afghanistan

5 = Albania

6 = Algeria

7 = Andorra

8 = Angola

10 = Antigua and Barbuda

11 = Argentina

12 = Armenia

14 = Australia

15 = Austria

16 = Azerbaijan

17 = Bahamas

18 = Bahrain

19 = Bangladesh

20 = Barbados

21 = Belgium

22 = Belize

23 = Benin

24 = Bermuda*

25 = Bhutan

26 = Bolivia

28 = Bosnia-Herzegovina

29 = Botswana

30 = Brazil

31 = Brunei

32 = Bulgaria

33 = Burkina Faso

34 = Burundi

35 = Belarus

36 = Cambodia

37 = Cameroon

38 = Canada

40 = Cayman Islands

41 = Central African Republic

42 = Chad

43 = Chile

44 = China

45 = Colombia

46 = Comoros

47 = Republic of the Congo

49 = Costa Rica

50 = Croatia

51 = Cuba

53 = Cyprus

54 = Czech Republic

55 = Denmark

56 = Djibouti

57 = Dominica

58 = Dominican Republic

59 = Ecuador

60 = Egypt

61 = El Salvador

62 = Equatorial Guinea

63 = Eritrea

64 = Estonia

65 = Ethiopia

66 = Falkland Islands

67 = Fiji

68 = Finland

69 = France

70 = French Guiana

71 = French Polynesia

72 = Gabon

73 = Gambia

74 = Georgia

75 = Germany

76 = Ghana

77 = Gibraltar

78 = Greece

79 = Greenland*

80 = Grenada

81 = Guadeloupe

83 = Guatemala

84 = Guinea

85 = Guinea-Bissau

86 = Guyana

87 = Haiti

88 = Honduras

89 = Hong Kong

90 = Hungary

91 = Iceland

92 = India

93 = Indonesia

94 = Iran

95 = Iraq

96 = Ireland

97 = Israel

98 = Italy

99 = Ivory Coast

100 = Jamaica

101 = Japan

102 = Jordan

103 = Kazakhstan

104 = Kenya

106 = Kuwait

107 = Kyrgyzstan

108 = Laos

109 = Latvia

110 = Lebanon

111 = Lesotho

112 = Liberia

113 = Libya

114 = Liechtenstein*

115 = Lithuania

116 = Luxembourg

117 = Macau

118 = Macedonia

119 = Madagascar

120 = Malawi

121 = Malaysia

122 = Maldives

123 = Mali

124 = Malta

125 = Man, Isle of*

127 = Martinique

128 = Mauritania

129 = Mauritius

130 = Mexico

132 = Moldova

134 = Mongolia*

136 = Morocco

137 = Mozambique

138 = Myanmar

139 = Namibia

141 = Nepal

142 = Netherlands

143 = New Caledonia

144 = New Zealand

145 = Nicaragua

146 = Niger

147 = Nigeria

149 = North Korea

151 = Norway

152 = Oman*

153 = Pakistan

155 = West Bank and Gaza Strip

156 = Panama

157 = Papua New Guinea

158 = Paraguay

159 = Peru

160 = Philippines

161 = Poland

162 = Portugal

163 = Puerto Rico*

164 = Qatar

166 = Romania

167 = Russia

168 = Rwanda

173 = Saudi Arabia

174 = Senegal

175 = Serbia-Montenegro

176 = Seychelles

177 = Sierra Leone

178 = Singapore

179 = Slovak Republic

180 = Slovenia

181 = Solomon Islands

182 = Somalia

183 = South Africa

184 = South Korea

185 = Spain

186 = Sri Lanka

189 = St. Kitts and Nevis

190 = St. Lucia

192 = St. Martin*

195 = Sudan

196 = Suriname

197 = Swaziland

198 = Sweden

199 = Switzerland

200 = Syria

201 = Taiwan

202 = Tajikistan

203 = Tanzania

204 = Togo

205 = Thailand

206 = Tonga*

207 = Trinidad and Tobago

208 = Tunisia

209 = Turkey

210 = Turkmenistan

213 = Uganda

214 = Ukraine

215 = United Arab Emirates

216 = Great Britain*

217 = United States

218 = Uruguay

219 = Uzbekistan

220 = Vanuatu

221 = Vatican City

222 = Venezuela

223 = Vietnam

226 = Wallis and Futuna

228 = Yemen

229 = Democratic Republic of the Congo

230 = Zambia

231 = Zimbabwe

233 = Northern Ireland*

235 = Yugoslavia

236 = Czechoslovakia

238 = Corsica*

334 = Asian*

347 = East Timor

349 = Western Sahara

351 = Commonwealth of Independent States*

359 = Soviet Union

362 = West Germany (FRG)

377 = North Yemen

403 = Rhodesia

406 = South Yemen

422 = International

428 = South Vietnam

499 = East Germany (GDR)

520 = Sinhalese*

532 = New Hebrides

603 = United Kingdom

604 = Zaire

605 = People's Republic of the Congo

999 = Multinational*

1001 = Serbia

1002 = Montenegro

1003 = Kosovo

1004 = South Sudan

5. Region

(region)

Categorical Variable

This field identifies the region in which the incident occurred. The regions are divided into the following 12 categories, and dependent on the country coded for the case:

1 = North America

Canada, Mexico, United States

2 = Central America & Caribbean

Antigua and Barbuda, Bahamas, Barbados, Belize, Cayman Islands, Costa Rica, Cuba, Dominica, Dominican Republic, El Salvador, Grenada, Guadeloupe, Guatemala, Haiti, Honduras, Jamaica, Martinique, Nicaragua, Panama, St. Kitts and Nevis, St. Lucia, Trinidad and Tobago

3 = South America

Argentina, Bolivia, Brazil, Chile, Colombia, Ecuador, Falkland Islands, French Guiana, Guyana, Paraguay, Peru, Suriname, Uruguay, Venezuela

4 = East Asia

China, Hong Kong, Japan, Macau, North Korea, South Korea, Taiwan

5 = Southeast Asia

Brunei, Cambodia, East Timor, Indonesia, Laos, Malaysia, Myanmar, Philippines, Singapore, South Vietnam, Thailand, Vietnam

6 = South Asia

Afghanistan, Bangladesh, Bhutan, India, Maldives, Mauritius, Nepal, Pakistan, Sri Lanka

7 = Central Asia

Armenia, Azerbaijan, Georgia, Kazakhstan, Kyrgyzstan, Tajikistan, Turkmenistan, Uzbekistan

8 = Western Europe

Andorra, Austria, Belgium, Cyprus, Denmark, Finland, France, Germany, Gibraltar, Greece, Iceland, Ireland, Italy, Luxembourg, Malta, Netherlands, Norway, Portugal, Spain, Sweden, Switzerland, United Kingdom, Vatican City, West Germany (FRG)

9 = Eastern Europe

Albania, Belarus, Bosnia-Herzegovina, Bulgaria, Croatia, Czech Republic, Czechoslovakia, East Germany (GDR), Estonia, Hungary, Kosovo, Latvia, Lithuania, Macedonia, Moldova, Montenegro, Poland, Romania, Russia, Serbia, Serbia-Montenegro, Slovak Republic, Slovenia, Soviet Union, Ukraine, Yugoslavia

10 = Middle East & North Africa

Algeria, Bahrain, Egypt, Iran, Iraq, Israel, Jordan, Kuwait, Lebanon, Libya, Morocco, North Yemen, Qatar, Saudi Arabia, South Yemen, Syria, Tunisia, Turkey, United Arab Emirates, West Bank and Gaza Strip, Western Sahara, Yemen

11 = Sub-Saharan Africa

Angola, Benin, Botswana, Burkina Faso, Burundi, Cameroon, Central African Republic, Chad, Comoros, Democratic Republic of the Congo, Djibouti, Equatorial Guinea, Eritrea, Ethiopia, Gabon, Gambia, Ghana, Guinea, Guinea-Bissau, Ivory Coast, Kenya, Lesotho, Liberia, Madagascar, Malawi, Mali, Mauritania, Mozambique, Namibia, Niger, Nigeria, People's Republic of the Congo, Republic of the Congo, Rhodesia, Rwanda, Senegal, Seychelles, Sierra Leone, Somalia, South Africa, South Sudan, Sudan, Swaziland, Tanzania, Togo, Uganda, Zaire, Zambia, Zimbabwe

12 = Australasia & Oceania

Australia, Fiji, French Polynesia, New Caledonia, New Hebrides, New Zealand, Papua New Guinea, Solomon Islands, Vanuatu, Wallis and Futuna

6. Extended Incident

(extended)

Categorical Variable

1 = "Yes" The duration of an incident extended more than 24 hours.

0 = "No" The duration of an incident extended less than 24 hours.

7. Vicinity

(vicinity)

Categorical Variable

1 = "Yes" The incident occurred in the immediate vicinity of the city in question.

0 = "No" The incident in the city itself.

8. Inclusion Criteria

(crit1, crit2, crit3)

Categorical Variables

These variables record which of the inclusion criteria (in addition to the necessary (criteria) are met.

Criterion 1: POLITICAL, ECONOMIC, RELIGIOUS, OR SOCIAL GOAL (CRIT1)

The violent act must be aimed at attaining a political, economic, religious, or social goal. This criterion is not satisfied in those cases where the perpetrator(s) acted out of a pure profit motive or from an idiosyncratic personal motive unconnected with broader societal change.

1 = "Yes" The incident meets Criterion 1.

0 = "No" The incident does not meet Criterion 1 or there is no indication that the incident meets Criterion 1.

Criterion 2: INTENTION TO COERCE, INTIMIDATE OR PUBLICIZE TO LARGER AUDIENCE(S) (CRIT2)

To satisfy this criterion there must be evidence of an intention to coerce, intimidate, or convey some other message to a larger audience (or audiences) than the immediate victims. Such evidence can include (but is not limited to) the following: pre- or post-attack statements by the perpetrator(s), past behavior by the perpetrators, or the particular nature of the target/victim, weapon, or attack type.

1 = "Yes" The incident meets Criterion 2.

0 = "No" The incident does not meet Criterion 2 or no indication.

Criterion 3: OUTSIDE INTERNATIONAL HUMANITARIAN LAW (CRIT3)

The action is outside the context of legitimate warfare activities, insofar as it targets non-combatants (i.e. the act must be outside the parameters permitted by international humanitarian law as reflected in the Additional Protocol to the Geneva Conventions of 12 August 1949 and elsewhere).

1 = "Yes" The incident meets Criterion 3.

0 = "No" The incident does not meet Criterion 3.

9. Part of Multiple Incident

(multiple)

Categorical Variable

In those cases where several attacks are connected, but where the various actions do not constitute a single incident (either the time of occurrence of incidents or their locations are discontinuous – see Single Incident Determination section above), then "Yes" is selected to denote that the particular attack was part of a "multiple" incident.

1 = "Yes" The attack is part of a multiple incident.

0 = "No" The attack is not part of a multiple incident.

Note: This field is presently only systematically available with incidents occurring after 1997.

10.Successful Attack

(success)

Categorical Variable

Success of a terrorist strike is defined according to the tangible effects of the attack. Success is not judged in terms of the larger goals of the perpetrators. For example, a bomb that exploded in a building would be counted as a success even if it did not succeed in bringing the building down or inducing government repression. The definition of a successful attack depends on the type of attack. Essentially, the key question is whether or not the attack type took place. If a case has multiple attack types, it is successful if any of the attack types are successful, with the exception of assassinations, which are only successful if the intended target is killed.

1 = "Yes" The incident was successful.

0 = "No" The incident was not successful.

11.Suicide Attack

(suicide)

Categorical Variable

This variable is coded "Yes" in those cases where there is evidence that the perpetrator did not intend to escape from the attack alive.

1 = "Yes" The incident was a suicide attack.

0 = "No" There is no indication that the incident was a suicide attack.

12.Attack Type

(attacktype1)

Categorical Variable

This field captures the general method of attack and often reflects the broad class of tactics used. It consists of nine categories, which are defined below. Up to three attack types can be recorded for each incident. Typically, only one attack type is recorded for each incident unless the attack is comprised of a sequence of events. When multiple attack types may apply, the most appropriate value is determined based on the hierarchy below. For example, if an assassination is carried out through the use of an explosive, the Attack Type is coded as Assassination, not Bombing/Explosion. If an attack involves a sequence of events, then the first, the second, and the third attack types are coded in the order of the hierarchy below rather than the order in which they occurred.

Attack Type Hierarchy:

Assassination

Hijacking

Kidnapping

Barricade Incident

Bombing/Explosion

Unknown

Armed Assault

Unarmed Assault

Facility/Infrastructure Attack

1 = Assassination

An act whose primary objective is to kill one or more specific, prominent individuals. Usually carried out on persons of some note, such as high-ranking military officers, government officials, celebrities, etc. Not to include attacks on non-specific members of a targeted group. The killing of a police officer would be an armed assault unless there is reason to believe the attackers singled out a particularly prominent officer for assassination.

2 = Armed Assault

An attack whose primary objective is to cause physical harm or death directly to human beings by use of a firearm, incendiary, or sharp instrument (knife, etc.). Not to include attacks involving the use of fists, rocks, sticks, or other handheld (less-than-lethal) weapons. Also includes attacks involving certain classes of explosive devices in addition to firearms, incendiaries, or sharp instruments. The explosive device subcategories that are included in this classification are grenades, projectiles, and unknown or other explosive devices that are thrown.

3 = Bombing/Explosion

An attack where the primary effects are caused by an energetically unstable material undergoing rapid decomposition and releasing a pressure wave that causes physical damage to the surrounding environment. Can include either high or low explosives (including a dirty bomb) but does not include a nuclear explosive device that releases energy from fission and/or fusion, or an incendiary device where decomposition takes place at a much slower rate. If an attack involves certain classes of explosive devices along with firearms, incendiaries, or sharp objects, then the attack is coded as an armed assault only. The explosive device subcategories that are included in this classification are grenades, projectiles, and unknown or other explosive devices that are thrown in which the bombers are also using firearms or incendiary devices.

4 = Hijacking

An act whose primary objective is to take control of a vehicle such as an aircraft, boat, bus, etc. for the purpose of diverting it to an unprogrammed destination, force the release of prisoners, or some other political objective. Obtaining payment of a ransom should not the sole purpose of a Hijacking, but can be one element of the incident so long as additional objectives have also been stated. Hijackings are distinct from Hostage Taking because the target is a vehicle, regardless of whether there are people/passengers in the vehicle.

5 = Hostage Taking (Barricade Incident)

An act whose primary objective is to take control of hostages for the purpose of achieving a political objective through concessions or through disruption of normal operations. Such attacks are distinguished from kidnapping since the incident occurs and usually plays out at the target location with little or no intention to hold the hostages for an extended period in a separate clandestine location.

6 = Hostage Taking (Kidnapping)

An act whose primary objective is to take control of hostages for the purpose of achieving a political objective through concessions or through disruption of normal operations. Kidnappings are distinguished from Barricade Incidents (above) in that they involve moving and holding the hostages in another location.

7 = Facility/ Infrastructure Attack

An act, excluding the use of an explosive, whose primary objective is to cause damage to a non-human target, such as a building, monument, train, pipeline, etc. Such attacks include arson and various forms of sabotage (e.g., sabotaging a train track is a facility/infrastructure attack, even if passengers are killed). Facility/infrastructure attacks can include acts which aim to harm an installation, yet also cause harm to people incidentally (e.g. an arson attack primarily aimed at damaging a building, but causes injuries or fatalities).

8 = Unarmed Assault

An attack whose primary objective is to cause physical harm or death directly to human beings by any means other than explosive, firearm, incendiary, or sharp instrument (knife, etc.). Attacks involving chemical, biological or radiological weapons are considered unarmed assaults.

9 = Unknown

The attack type cannot be determined from the available information.

13.Target/Victim Type

(targtype1)

Categorical Variable

The target/victim type field captures the general type of target/victim. When a victim is attacked specifically because of his or her relationship to a particular person, such as a prominent figure, the target type reflects that motive. For example, if a family member of a government official is attacked because of his or her relationship to that individual, the type of target is “government.” This variable consists of the following 22 categories:

1 = Business

Businesses are defined as individuals or organizations engaged in commercial or mercantile activity as a means of livelihood. Any attack on a business or private citizens patronizing a business such as a restaurant, gas station, music store, bar, café, etc. This includes attacks carried out against corporate offices or employees of firms like mining companies, or oil corporations. Furthermore, includes attacks conducted on business people or corporate officers. Included in this value as well are hospitals and chambers of commerce and cooperatives. Does not include attacks carried out in public or quasi-public areas such as "business district or commercial area", or generic business-related individuals such as "businessmen" (these attacks are captured under "Private Citizens and Property", see below.) Also does not include attacks against generic business-related individuals such as "businessmen." Unless the victims were targeted because of their specific business affiliation, these attacks belong in "Private Citizens and Property."

2 = Government (general)

Any attack on a government building; government member, former members, including members of political parties in official capacities, their convoys, or events sponsored by political parties; political movements; or a government sponsored institution where the attack is expressly carried out to harm the government. This value includes attacks on judges, public attorneys (e.g., prosecutors), courts and court systems, politicians, royalty, head of state, government employees (unless police or military), election-related attacks, or intelligence agencies and spies. This value does not include attacks on political candidates for office or members of political parties that do not hold an elected office (these attacks are captured in "Private Citizens and Property").

3 = Police

This value includes attacks on members of the police force or police installations; this includes police boxes, patrols headquarters, academies, cars, checkpoints, etc. Includes attacks against jails or prison facilities, or jail or prison staff or guards.

4 = Military

Includes attacks against military units, patrols, barracks, convoys, jeeps, and aircraft. Also includes attacks on recruiting sites, and soldiers engaged in internal policing functions such as at checkpoints and in anti-narcotics activities. This category also includes peacekeeping units that conduct military operations (e.g., AMISOM) Excludes attacks against non-state militias and guerrillas, these types of attacks are coded as "Terrorist/Non-state Militias" see below.

5 = Abortion Related

Attacks on abortion clinics, employees, patrons, or security personnel stationed at clinics.

6 = Airports & Aircraft

An attack that was carried out either against an aircraft or against an airport. Attacks against airline employees while on board are also included in this value. Includes attacks conducted against airport business offices and executives. Military aircraft are not included.

7 = Government (Diplomatic)

Attacks carried out against foreign missions, including embassies, consulates, etc. This value includes cultural centers that have diplomatic functions, and attacks against diplomatic staff and their families (when the relationship is relevant to the motive of the attack) and property. The United Nations is a diplomatic target.

8 = Educational Institution

Attacks against schools, teachers, or guards protecting school sites. Includes attacks against university professors, teaching staff and school buses. Moreover, includes attacks against religious schools in this value. As noted below in the "Private Citizens and Property" value, the GTD has several attacks against students. If attacks involving students are not expressly against a school, university or other educational institution or are carried out in an educational setting, they are coded as private citizens and property. Excludes attacks against military schools (attacks on military schools are coded as "Military," see below).

9 = Food or Water Supply

Attacks on food or water supplies or reserves are included in this value. This generally includes attacks aimed at the infrastructure related to food and water for human consumption.

10 = Journalists & Media

Includes, attacks on reporters, news assistants, photographers, publishers, as well as attacks on media headquarters and offices. Attacks on transmission facilities such as antennae or transmission towers, or broadcast infrastructure are coded as "Telecommunications," see below.

11 = Maritime (Includes ports and maritime facilities)

Includes civilian maritime: attacks against fishing ships, oil tankers, ferries, yachts, etc. (Attacks on fishermen are coded as "Private Citizens and Property," see below).

12 = NGO

Includes attacks on offices and employees of non-governmental organizations (NGOs). NGOs here include large multinational non-governmental organizations such as the Red Cross and Doctors without Borders, as well as domestic organizations. Does not include labor unions, social clubs, student groups, and other non-NGO (such cases are coded as “Private Citizens and Property”, see below).

13= Other

This value includes acts of terrorism committed against targets which do not fit into other categories. Some examples include ambulances, firefighters, refugee camps, and international demilitarized zones.

14= Private Citizens & Property

This value includes attacks on individuals, the public in general or attacks in public areas including markets, commercial streets, busy intersections and pedestrian malls. Also includes ambiguous cases where the target/victim was a named individual, or where the target/victim of an attack could be identified by name, age, occupation, gender or nationality. This value also includes ceremonial events, such as weddings and funerals. The GTD contains a number of attacks against students. If these attacks are not expressly against a school, university or other educational institution or are not carried out in an educational setting, these attacks are coded using this value. Also, includes incidents involving political supporters as private citizens and property, provided that these supporters are not part of a government-sponsored event. Finally, this value includes police informers. Does not include attacks causing civilian casualties in businesses such as restaurants, cafes or movie theaters (these categories are coded as “Business” see above).

15 = Religious Figures/Institutions

This value includes attacks on religious leaders, (Imams, priests, bishops, etc.), religious institutions (mosques, churches), religious places or objects (shrines, relics, etc.). This value also includes attacks on organizations that are affiliated with religious entities that are not NGOs, businesses or schools. Attacks on religious pilgrims are considered “Private Citizens and Property;” attacks on missionaries are considered religious figures.

16 = Telecommunication

This includes attacks on facilities and infrastructure for the transmission of information. More specifically this value includes things like cell phone towers, telephone booths, television transmitters, radio, and microwave towers.

17 = Terrorists /Non-State Militias

Terrorists or members of identified terrorist groups within the GTD are included in this value. Membership is broadly defined and includes informants for terrorist groups, but excludes former or surrendered terrorists.

This value also includes cases involving the targeting of militias and guerillas.

18 = Tourists

This value includes the targeting of tour buses, tourists, or “tours.” Tourists are persons who travel primarily for the purposes of leisure or amusement. Government tourist offices are included in this value. The attack must clearly target tourists, not just an assault on a business or transportation system used by tourists. Travel agencies are coded as business targets.

19 = Transportation (other than aviation)

Attacks on public transportation systems are included in this value. This can include efforts to assault public buses, minibuses, trains, metro/subways, highways (if the highway itself is the target of the attack), bridges, roads, etc. The GTD contains a number of attacks on generic terms such as “cars” or “vehicles.” These attacks are assumed to be against “Private Citizens and Property” unless shown to be against public transportation systems. In this regard, buses are assumed to be public transportation unless otherwise noted.

20 = Unknown

The target type cannot be determined from the available information.

21 = Utilities

This value pertains to facilities for the transmission or generation of energy. For example, power lines, oil pipelines, electrical transformers, high tension lines, gas and electric substations, are all included in this value. This value also includes lampposts or street lights. Attacks on officers, employees or facilities of utility companies excluding the type of facilities above are coded as business.

22 = Violent Political Parties

This value pertains to entities that are both political parties (and thus, coded as “government” in this coding scheme) and terrorists. It is operationally defined as groups that engage in electoral politics and appear as “Perpetrators” in the GTD.

14.Target/Victim Subtype

(targsubtype1)

Categorical Variable

The target subtype variable captures the more specific target category and provides the next level of designation for each target type. If a target subtype is not applicable this variable is left blank.The subtypes for each target type are as follows:

Business

1 = Gas/Oil

2 = Restaurant/Bar/Café

3 = Bank/Commerce

4 = Multinational Corporation

5 = Industrial/Textiles/Factory

6 = Medical/Pharmaceutical

7 = Retail/Grocery/Bakery (including cell phone shops and generic shops)

8 = Hotel/Resort

9 = Farm/Ranch

10 = Mining

11 = Entertainment/Cultural/Stadium/Casino

12 = Construction

13 = Private Security Company/Firm

Government (General)

14 = Judges/Attorneys/Courts

15 = Politician or Political Party Movement/Meeting/Rally

16 = Royalty

17 = Head of State

18 = Government Personnel (excluding police, military)

19 = Election-related

20 = Intelligence

21 = Government Buildings/Facility/Office

Police

22 = Police Buildings (Headquarters/Stations/School)

23 = Police Patrol (including vehicles and convoys)

24 = Police Checkpoint

25 = Police Security Forces/Officers

26 = Prison/Jail

Military

27 = Military Barracks/Base/Headquarters/Checkpost

28 = Military Recruiting Station/Academy

29 = Military Unit/Patrol/Convoy

30 = Navy

31 = Air

32 = Coast Guard

33 = Army

34 = Military Personnel (soldiers, troops, officers, forces)

35 = Military Transportation/Vehicle (excluding convoys)

36 = Military Checkpoint

37 = North Atlantic Treaty Organization (NATO) Related

38 = Marine

39 = Paramilitary

Abortion Related

40 = Clinics

41 = Personnel

Airports & Aircraft

42 = Aircraft (not at an airport)

43 = Airline Officer/Personnel

44 = Airport

Government (Diplomatic)

45 = Diplomatic Personnel (outside of embassy, consulate)

46 = Embassy/Consulate

47 = International Organization (peacekeeper, aid agency, compound)

Educational Institution

48 = Teacher/Professor/Instructor

49 = School/University/Educational Building

50 = Other Personnel

Food and Water Supply

51 = Food Supply

52 = Water Supply

Journalists & Media

53 = Newspaper Journalist/Staff/Facility

54 = Radio Journalist/Staff/Facility

55 = Television Journalist/Staff/Facility

56 = Other (including online news agencies)

Maritime

57 = Civilian Maritime

58 = Commercial Maritime

59 = Oil Tanker

60 = Port

NGO

61 = Domestic NGO

62 = International NGO

Other

63 = Ambulance

64 = Fire Fighter/Truck

65 = Refugee Camp

66 = Demilitarized Zone (including Green Zone)

Private Citizens & Property

67 = Unnamed Civilian/Unspecified

68 = Named Civilian

69 = Religion Identified

70 = Student

71 = Race/Ethnicity Identified

72 = Farmer

73 = Vehicles/Transportation

74 = Marketplace/Plaza/Square

75 = Village/City/Town/Suburb

76 = House/Apartment/Residence

77 = Laborer (General)/Occupation Identified

78 = Procession/Gathering (funeral, wedding, birthday, religious)

79 = Public Areas (e.g., Public garden, parking lot, garage, beach, public buildings, camps)

80 = Memorial/Cemetery/Monument

81 = Museum/Cultural Center/Cultural House

82 = Labor Union Related

83 = Protester

84 = Political Party Member/Rally

Religious Figures/Institutions

85 = Religious Figure

86 = Place of Worship

87 = Affiliated Institution

Telecommunication

88 = Radio

89 = Television

90 = Telephone/Telegraph

91 = Internet Infrastructure

92 = Multiple Telecommunication Targets

Terrorist/Non-State Militia

93 = Terrorist Organization

94 = Non-State Militia

Tourists

95 = Tourism Travel Agency

96 = Tour Bus/Van/Vehicle

97 = Tourist

98 = Other Facility

Transportation

99 = Bus (excluding tourist)

100 = Train/Train Tracks/ Trolley

101 = Bus Station/Stop

102 = Subway

103 = Bridge/Car Tunnel

104 = Highway/Road/Toll/Traffic Signal

105 = Taxi/Rickshaw

Unknown

[No corresponding target subtypes]

Utilities

106 = Gas

107 = Electricity

108 = Oil

Violent Political Parties

109 = Party Official/Candidate/Other Personnel

110 = Party Office/Facility

111 = Rally

15.Nationality of Target/Victim

(natlty1)

Categorical Variable

This is the nationality of the target that was attacked, and is not necessarily the same as the country in which the incident occurred, although in most cases it is. For hijacking incidents, the nationality of the plane is recorded and not that of the passengers. For numeric nationality codes, please see the country codes

16.First Perpetrator Group Suspected/Unconfirmed

(guncertain1)

Categorical Variable

This variable indicates whether or not the information reported by sources about the Perpetrator Group Name(s) is based on speculation or dubious claims of responsibility.

1 = "Yes" The perpetrator attribution(s) for the incident are suspected.

0 = "No" The perpetrator attribution(s) for the incident are not suspected.

17.Weapon Type

(weaptype1)

Categorical Variable

Up to four weapon types are recorded for each incident. This field records the general type of weapon used in the incident. It consists of the following 13 categories:

1 = Biological

A weapon whose components are produced from pathogenic microorganisms or toxic substances of biological origins.

2 = Chemical

A weapon produced from toxic chemicals that is contained in a delivery system and dispersed as a liquid, vapor, or aerosol.

3 = Radiological

A weapon whose components are produced from radioactive materials that emit ionizing radiation and can take many forms.

4 = Nuclear

A weapon which draws its explosive force from fission, fusion, or a combination of these methods.

5 = Firearms

A weapon which is capable of firing a projectile using an explosive charge as a propellant.

6 = Explosives/Bombs/Dynamite

A weapon composed of energetically unstable material undergoing rapid decomposition and releasing a pressure wave that causes physical damage to the surrounding environment.

7 = Fake Weapons

A weapon that was claimed by the perpetrator at the time of the incident to be real but was discovered after-the-fact to be non-existent or incapable of producing the desired effects.

8 = Incendiary

A weapon that is capable of catching fire, causing fire, or burning readily and produces intensely hot fire when exploded.

9 = Melee

A weapon—targeting people rather than property—that does not involve a projectile in which the user and target are in contact with it simultaneously.

10 = Vehicle

An automobile that is used in an incident that does not incorporate the use of explosives such as a car bomb or truck bomb.

11 = Sabotage Equipment

A weapon that is used in the demolition or destruction of property (e.g., removing bolts from a train tracks).

12 = Other

A weapon that has been identified but does not fit into one of the above categories.

13 = Unknown

The weapon type cannot be determined from the available information.

18.Total Number of Fatalities

(nkill)

Numeric Variable

This field stores the number of total confirmed fatalities for the incident. The number includes all victims and attackers who died as a direct result of the incident. Where there is evidence of fatalities, but a figure is not reported or it is too vague to be of use, this field remains blank. If information is missing regarding the number of victims killed in an attack, but perpetrator fatalities are known, this value will reflect only the number of perpetrators who died as a result of the incident. Likewise, if information on the number of perpetrators killed in an attack is missing, but victim fatalities are known, this field will only report the number of victims killed in the incident.

19.Total Number of Injured

(nwound)

Numeric Variable

This field records the number of confirmed non-fatal injuries to both perpetrators and victims. It follows the conventions of the “Total Number of Fatalities” field described above.

20.Hostages or Kidnapping Victims

(ishostkid)

Categorical Variable

This field records whether or not the victims were taken hostage (i.e. held against their will) or kidnapped (i.e. held against their will and taken to another location) during an incident.

1 = "Yes" The victims were taken hostage or kidnapped.

0 = "No" The victims were not taken hostage or kidnapped.

-9 = "Unknown" It is unknown if the victims were taken hostage or kidnapped.

21.International- Miscellaneous

(INT_MISC)

(Categorical Variable)

This variable is based on a comparison between the location of the attack and the nationality of the target(s)/victim(s). It indicates whether a perpetrator group attacked a target of a different nationality.

Unlike the logistically international and ideologically international variables, it does not require information about the nationality of the perpetrator group. If an attack is international on this dimension, it is necessarily also either logistically international or ideologically international, but it is not clear which one. If an attack is domestic on this dimension, it may also be logistically international or ideologically international, or domestic on all dimensions.

1 = "Yes" The attack was miscellaneous international; the location of the attack differs from the nationality of the target(s)/victim(s).

0 = "No" The attack was miscellaneous domestic; the location of the attack is the same as the nationalities of the target(s)/victim(s).

-9 = "Unknown" It is unknown if the attack was miscellaneous international or domestic; the nationality of target/victim is unknown.