Ryerson University
Student: Emil Ibrahim
Student Id: 500862456

Date: 13-Feb-2019
Course: CKM136XJ0 Capstone
Supervisor: Dr. Can Kavaklioglu

Global Terrorism Data Analysis

Introduction:

Terrorism is one of main problems in the 21st century.
Terrorism has continuously been an ongoing danger all over the world.
Analysing global terrorism historical dataset will help to understand the insights of the problem and be prepared in future to prevent or mitigate the number or impact of attacks.

Context:

Information on more than 180,000 Terrorist Attacks.
The Global Terrorism Database (GTD) is an open-source database including information on terrorist attacks around the world from 1970 through 2017. The GTD includes systematic data on domestic as well as international terrorist incidents that have occurred during this time period and now includes more than 180,000 attacks.

Content:

Geography: Worldwide.
Time period: 1970-2017, *except 1993.*
Variables: >135 variables on location, tactics, perpetrators, targets, and outcomes ,….etc

Research question(s):

Can the success of an attack and number of casualties (killed and wounded) be predicted by knowing attack features?

Literature Review:

Suspicious human activity recognition [1]:
From surveillance video this paper used image processing and computer vision to monitor the human activities in real-time and categorize them as usual and unusual activities.
This paper consists of six abnormal activities such as abandoned object detection, theft detection, fall detection, accidents and illegal parking detection on road, violence activity detection, and fire detection.
The classifier used various models such as SVM, Random forest, K-NN.

The Psychology of Suicide Terrorism [2]:
The paper reviewed the understanding of the psychology of suicide terrorism, the study shows that the outstanding common characteristics of a terrorists is their normality.
The paper examines various psychological models that may contribute to the expansion of suicide bombing.

Suicide Terrorists: Are They Suicidal? [3]

The study found many discrepancies uncovered between suicide terrorists and other suicides.

The paper suggest that the terrorists are not truly suicidal and should not be viewed as a subgroup of the general suicide population.

The paper also covered the motivations for suicide vs suicide terrorism.

Terrorism and the internet: a double-edged sword [4]:

The purpose of this paper was to highlight the extent of the use of the internet by terrorist organisations to achieve their strategic and operational objectives.

The study showed that one significant enabler has been the internet, which enables anonymous communication, aids recruitment, encourages the sharing of knowledge, as well as playing a significant role in the spreading of propaganda.

The understanding of the use of internet-based technologies, not only as a potential target for terrorist attack, but as a tool to achieve its ideological and operational goals, remains a neglected area of study requiring further focus.

The paper focused on Al-Qaeda's use of the internet to plan operations for 9/11 was brought to light when computers seized in Afghanistan revealed that al-Qaeda was collecting intelligence on targets and sending encrypted messages via the internet, as well as using internet-based phones. Further investigation revealed the existence of various al-Qaeda web sites.

PATHWAY TO JIHADIST TERRORISM [5]: A CRITICAL LITERATURE REVIEW

The paper addressed a comparison of 10 studies, it identifies commonly seen characteristics of individuals engaged in neo-jihadist terrorism in the United States.

Due to the diverse demographic factors of the individuals that engage in jihad, it is difficult to identify who will become radicalized and engage in acts of terrorism. The number of foreign fighters has grown through recruitment largely from al-Qaeda and ISIS. The third-wave of jihad has also introduced a new threat from homegrown extremists through self-radicalization.

Terrorism Risk Forecasting by Ideology [6]:

The paper presented a study to produce foresights on terrorism.

The paper suggests that target selection and the places where terrorist attacks occur are related to a group's strategies, and violent terrorist acts vary with respect to their surroundings in the same jurisdiction when the ideology is the point of comparison.

The analysis begins with a comparison of targets and risk factors by the ideology of the perpetrators.

After identifying high-risk locations, tested the predictive validity of risk terrain modeling.

The study results showed that context and influence the risky areas of terrorism vary by the nature of the ideology in the jurisdiction.

Terrorists have always added geography into the equation to make a rational choice of their attacks, and the choice-making is ideology-dependent

The predictive model was done using logistic regression and used risk terrain modeling (RTM) to explore the relationship between terrorism, ideology, and geography.

Explaining religious terrorism [7]: A data-mined analysis

The paper examined the relationship between religious liberty and faith-based terrorism by looking specifically at terrorism motivated by a religious imperative and a country's level of religious liberty.

Using classification data mining, paper tested a unique dataset on religious terrorism in order to discover the characteristics that contribute to a country experiencing religiously motivated terrorism. The analysis finds that religious terrorism is indeed a product of a dearth of religious liberty.

The data mining process involves both human and software resources. Data mining is used not only to predict the outcome of a future event but also to provide knowledge about the structure and interrelationships between data. Classification mining algorithms are used to create models that describe existing data and relationships within that dataset. The resulting model is expressed as a classification tree.

This article has made a simple but important claim: the denial of religious freedom increases the likelihood of violent religious forms of political engagement; paradoxically, the best way to combat religious terrorism is not by restricting religious practices but rather by safeguarding their legitimate manifestations.

The analytical study of terrorism [8]: Taking stock

This article presents a review of the analytical study of terrorism that views all agents as rational decisionmakers.

The papers used the Global Terrorism Database (GTD), these include analyses of terrorist attack trends, the economic consequences of terrorism, the study of counterterrorism effectiveness, the causes of terrorism, and the relationship of terrorism and liberal democracies.

New developments in the field focused on distinguishing key differences between domestic and transnational terrorism. Other major developments involved the study of networked terrorists and the role of counterterrorism foreign aid.

Causality between terrorism and economic growth [9]:

This article analyzes the causal relationship between terrorism and economic growth, running a series of tests for a maximum of 160 countries from 1970 to 2007. The authors find that the causal relationship between terrorism and growth is heterogeneous over time and across space.

In order to examine the hypotheses regarding the terrorism–economy the data compiled on terrorism and economic growth for a maximum of 160 countries for the 1970–2007 period.

Exploring the relationship between global terrorist ideology and attack methodology [10]:

The paper used cluster analysis to test the hypothesis that a visual model could be developed to identify patterns in the last 45 years of global terrorism.

Terrorist attack methodologies of assassinations, armed assaults and bombings figured prominently in most of the segments, but national culture and ideology were also present as factors in some clusters.

Logistic regression was then applied to test the hypothesis that combinations of terrorist attack methodology and ideology could estimate the risk of a terrorist strike being successful.

The paper concluded that the odds ratio was statistically significant but the results were theoretically weak in supporting the hypothesis. Hijacking was 1.131 times more likely to be successful as compared with other types of global terrorist methodologies, and the relative risk of a terrorist strike being committed by foreign immigrants was 1.013 higher as compared with domestic terrorists.

Poverty, Political Freedom, and the Roots of Terrorism [11]:

The paper used a new dataset on terrorist risk world-wide, It failed to find a significant association between terrorism and economic variables such as income once the effect of other country characteristics is taken into account.

The estimates suggest that political freedom has a non-monotonic effect on terrorism. This result is consistent with the observed increase in terrorism for countries in transition from authoritarian regimes to democracies. The results also show that certain geographic characteristics may favor the presence of terrorist.

ISIL's Execution Videos: Audience Segmentation and Terrorist Communication in the Digital Age [12]:

This article offers a bottom-up understanding of the media strategy employed by the Islamic State of Iraq and the Levant (ISIL) as it relates to the production and dissemination of its hostage execution videos.

Through an empirical analysis of sixty-two videos of executions produced by ISIL in the year following its establishment as the "Islamic State" in 2014, this study examines the videos as a major component of ISIL's media strategy. Through these media products, ISIL seeks to spread a political message aimed at both local and global, ingroup and outgroup consumption through audience segmentation, while striving to influence both local and global audiences through the use and production of graphic violence. This article also discusses the strategy governing the production and release of ISIL's execution videos; how it relies on the global media to transmit its intertwined political and religious agenda in the digital media age.

The paper concluded that ISIL's use of execution video productions provides important context to the ways in which an emergent terrorist group frames its execution videos in terms of narrative, images, and ideology in order to set its political and religious agenda by relying on global and digital media to be spread and accessed. This empirical study has shed light on the complexity of the relationship between terrorism and the digital media age.

Terrorism Financing with Virtual Currencies [(13)]: Can Regulatory Technology Solutions Combat This?

This article considers the terrorism financing risk associated with the growth of Financial Technology innovations and in particular, focuses on virtual currency products and services. The ease with which cross-border payments by virtual currencies are facilitated, the anonymity surrounding their usage, and their potential to be converted into the fiat financial system, make them ideal for terrorism financing and therefore calls for a coordinated global regulatory response.

The paper concluded that the growth of the FinTech industry worldwide signals huge opportunities for businesses and consumers, it also introduces challenges to the global financial industry.
The challenge considered in this article is the potential of VC, a key FinTech innovation, to be used to finance terrorism activities.
This article has argued that, while financial regulatory regimes worldwide are at different stages of development and the robustness of law enforcement regimes worldwide vary, the threat of terrorism financing remains real—more so as terrorism financing is likely to shift to jurisdictions with weaker regimes.

Terrorism and Corruption [(14)]:
Alternatives for Goal Attainment Within Political Opportunity Structures

This study on the connection between corruption and political violence. It attempts to uncover whether domestic terrorism as a specific form of political violence.
The models required the measurement of the following key concepts:
political violence, perceived corruption, government coercion, regime repressiveness, cultural fractionalization, inequality, economic development, "heritage" of rebellion, transnational corporate penetration, level of education, and population.
The study used The Global Terrorism Database (GTD).
Analyses demonstrated that corruption and terrorist violence exhibit an inverse relationship, which supports the theoretical perspective that they are shared avenues within an extralegal opportunity structure, demonstrating that where the avenue of corruption has been restricted, countries experience greater rates of terrorist violence. This initial attempt to quantify the relationship between terrorist violence and corruption requires further analysis in order to identify under which conditions, in which regions, and over which time periods this relationship may vary.

Applying Analytical Methods to Study Terrorism [(15)]:

This paper study domestic and transnational terrorism insights.

For empirical applications, the paper focuses on the study of trend, cycles, and forecasting.
For forecasting Terrorism is to relate statistically the current number of terrorist incidents to their past values, time, and potential shocks or interventions (events such as a policy change or the rise in fundamentalism). There are some important forecasting insights. First, observed patterns in the time series can be used for forecasting purposes. Second, short-run forecasts are more accurate than long-run forecasts. Third, a time series of terrorist incidents can be forecast without knowing precisely why the numbers of incidents change in a particular regular pattern. Fourth, the ability to forecast patterns of, say, skyjackings does not mean that one can predict a particular catastrophic event such as 9 ⁄ 11.

PSYCHIATRY AND BEHAVIORAL SCIENCES [(16)]:
On the Radicalization Process

This study aimed to provide an in-depth description of the radicalization process. The author analysis is first based on the author's experience in the psychological evaluation of terrorist behavior and second on an exhaustive review of the current literature. The search terms "terrorism," "radicalization," "social psychology," and "psychopathology" were used to identify relevant studies.
The study performed with a focus on several aspects, such as radicalization risk factors, brainwashing, the role of the media, and finally, in de-radicalization programs.

The paper concluded that the radicalization process is an increasing and complex phenomenon. It implies several aspects that are multidimensional (on the individual and societal levels) and heterogeneous, such as some individual risk factors, the brainwashing, and cognitive modifications (i.e. the role of the media in general).

In terms of counter-radicalization programs, have shown mixed results. The most successful efforts in Britain have been the efforts of the so-called Channel program, which is part of the British government's counter-terrorism strategy, to divert young people from extremism. Such efforts, which involve the police, social services, and local authorities working together, draw on methods used to help young people leave gangs.

On a more practical note, it is clear there are currently not enough detailed case studies of terrorists to inform psychological analyses or even to conduct comprehensive reviews of the literature.

Dataset:

The Global Terrorism Database (GTD) is an open-source database including information on terrorist attacks around the world from 1970 through 2017. The GTD includes systematic data on domestic as well as international terrorist incidents that have occurred during this time period and now includes more than 180,000 attacks.

Content:

Geography: Worldwide.
Time period: 1970-2017, *except 1993.*
Variables: >135 variables on location, tactics, perpetrators, targets, and outcomes ,….etc

The source of the dataset:

https://www.kaggle.com/START-UMD/gtd#globalterrorismdb_0718dist.csv

Attributes:

GTD ID and Date

Attribute	Type
eventid	Numeric
iyear	Numeric
imonth	Numeric
iday	Numeric
approxdate	Text
extended	Categorical
resolution	Numeric Date

Incident Information

Attribute	Type
summary	Text
crit1	Categorical
crit2	Categorical
crit3	Categorical
doubtterr	Categorical
alternative	Categorical
alternative_txt	Categorical
multiple	Categorical
related	Text

Incident Location

Attribute	Type
country	Categorical
country_txt	Categorical
region	Categorical
region_txt	Categorical
provstate	Text
city	Text
vicinity	Categorical
location	Text
latitude	Numeric
longitude	Numeric
specificity	Categorical

Attack Information

Attribute	Type
attacktype1	Categorical
attacktype1_txt	Categorical
attacktype2	Categorical
attacktype2_txt	Categorical
attacktype3	Categorical
attacktype3_txt	Categorical
success	Categorical
suicide	Categorical

Additional Information & Sources

Attribute	Type
addnotes	Text
INT_LOG	Categorical
INT_IDEO	Categorical
INT_MISC	Categorical
INT_ANY	Categorical
scite1	Text
scite2	Text
scite3	Text
dbsource	Text

Weapon Information

Attribute	Type
weaptype1	Categorical
weaptype1_txt	Categorical
weapsubtype1	Categorical
weapsubtype1_txt	Categorical
weaptype2	Categorical
weaptype2_txt	Categorical
weapsubtype2	Categorical
weapsubtype2_txt	Categorical
weaptype3	Categorical
weaptype3_txt	Categorical
weapsubtype3	Categorical
weapsubtype3_txt	Categorical
weaptype4	Categorical
weaptype4_txt	Categorical
weapsubtype4	Categorical
weapsubtype4_txt	Categorical
weapdetail	Text

Target/Victim Information

Attribute	Type
targtype1	Categorical
targtype1_txt	Categorical
targsubtype1	Categorical
targsubtype1_txt	Categorical
corp1	Text
target1	Text
natlty1	Categorical
natlty1_txt	Categorical
targtype2	Categorical
targtype2_txt	Categorical
targsubtype2	Categorical
targsubtype2_txt	Categorical
corp2	Text
target2	Text
natlty2	Categorical
natlty2_txt	Categorical
targtype3	Categorical
targtype3_txt	Categorical
targsubtype3	Categorical
targsubtype3_txt	Categorical
corp3	Text
target3	Text
natlty3	Categorical
natlty3_txt	Categorical

Perpetrator Information

Attribute	Type
gname	Text
gsubname	Text
gname2	Text
gsubname2	Text
gname3	Text
gsubname3	Text
guncertain1	Categorical
guncertain2	Categorical
guncertain3	Categorical
individual	Categorical
nperps	Numerical
nperpcap	Numerical
claimed	Categorical
claimmode	Categorical
claimmode_txt	Categorical
compclaim	Categorical
claim2	Categorical
claimmode2	Categorical
claimmode2_txt	Categorical
claim3	Categorical
claimmode3	Categorical
claimmode3_txt	Categorical
motive	Text

Casualties and Consequences

Attribute	Type
nkill	Numeric
nkillus	Numeric
nkillter	Numeric
nwound	Numeric
nwoundus	Numeric
nwoundte	Numeric
property	Categorical
propextent	Categorical
propextent_txt	Categorical
propvalue	Numeric
propcomment	Text
ishostkid	Categorical
nhostkid	Numeric
nhostkidus	Numeric
nhours	Numeric
ndays	Numeric
divert	Text
kidhijcountry	Text
ransom	Categorical
ransomamt	Numeric
ransomamtus	Numeric
ransompaid	Numeric
ransompaidus	Numeric
ransomnote	Text
hostkidoutcome	Categorical
hostkidoutcome_txt	Categorical
nreleased	Numeric

Attributes Descriptions:

Please refer to the table in the link below:

(if the embedded link did not work the excel file is attached separately).

attributes.xlsx

Approach:

initial Analysis
- **Univariate Analysis** (detailed data dictionary / list all vaibles and understand what each varible mean and represent)(what is the source of each varible directly measured or calculated based on other varibles or does the vaible has a time domain assoicated) - (Decide on the dependent (target) varibale) (Assigning the corrcet data types and appropriate column names)(for numeric attributes check the 5 number summary min 1sQ mean 3rdQ max)(for categrical attributes check and decide on the levels , frequency) (check and deal with data inconsistencies , missing values , erros , duplicates , outliers (boxplots) , numeric signs , upper and lower cases , spaces or spcial charachters in strigs) (check distributions of the varibales : Norml distirbution?)(Low variance filter)(check the imbalance in the dependent varbalie) (check time varbiles) (Univariate visulaizations)
- **BiVariate Analysis** (pairwise relations) (pairwise visuliazations like scatter plots) (correlation analysis spearman or pearson).
- **Multivaritae Analysis** (relations between more than 2 varibales)(statistical tools such as one way analysis or rank or to compare the means.

Exploratory Analysis
- Normalizing / scaling
- Sub-setting the data (may foucs on one coutry or regin analysis or undersand the weapons used more ofen)
- Decision rules , association rules , n grams
- Clustring such as K-mean
- Hypotheise analysis (metntal inllness, education, GDP/economic,war status, unemployment ,internet growth / technlogy /telecom tech such as AlThuria phones, healthcare, weather, virtual currrencies such as cryptocoin, presedint of USA periods, UN head periods)
- NLP analysis for some text attributes

Dimensionality Reduction
- remove attributs with too many missing values
- remove attributis with zero or very low varince
- remove one of the attribits with high correlations with other - prefer the one with more missing values or lower varince
- Feature selection(decide on the importance of the attribute using statstical measures like information gain or Gini index) (Forward selection and backward elemination)
- Feature extraction (reduce dimonins by transformation PCA princibal complnet anaylsis or Linear dircrminate analysis)

Experimental Desgin
- Randamizing, Splitting the data into training and test sets (tranditional or Concept drift in case of time involved)
- Treatment for imblance (undersampling the majority class and oversampling the minority class)
- Cross validation such as 10 flods
- Bootstrapping resmapling

Modeling
- Classification vs regression
- linearity vs non linearty
- for classifier and adat are linear (SVM(linear kernel) , linear logistic regression , Naivebase)
- for calssifier and data are nonlinear (SVM(nonlinear kernel) , deceision tree ,random forest, KNN)
- Measure (distance measures: Euclidean or Manhattan, Similarity meausre: Cosine similarity)

Evaluation
- Regression problem (R squared , goodness of fit , AIC (Akaike information criterion) , RMSE (root mean square error)
- Classification (Confucion matric , ROC (receiver opecrating charchritices) , Accurracy , Recall , Prescision)
- compare performnce between diffrent models (contingency tables , multivariate analysis of variance)

Improving the Model
- deal with very high accuracy(Fixation variables , overfitting)
- deal with very low performance (many need new variables or more obervations)
- itterate to improve the model performance

Conclusions
- Infrences , dissucsuions
- Threates to validity (Internal , External , construct) and propose the soultion that may mitigate these threats
- Next/ future steps

Project Github link:
https://github.com/emilkaram/CKM136XJ0-Global-Terrorism-Data-Analytics-Capstone

References:

1. Rajesh Kumar Tripathi, Anand Singh Jalal and Subhash Chand Agrawal [Feb 2017]. **Suspicious human activity recognition.** Retrieved from:
    https://link-springer-com.ezproxy.lib.ryerson.ca/article/10.1007%2Fs10462-017-9545-7

2. Jerrold M; Ali, Farhana; Henderson, Schuyler W; Shanfield, Stephen; Victoroff, Jeff; [Spring 2009]. **The Psychology of Suicide Terrorism.** Retrieved from:
    https://search-proquest-com.ezproxy.lib.ryerson.ca/docview/220656477?pq-origsite=summon

3. Townsend, Ellen.**Suicide & Life - Threatening Behavior; Washington** Vol. 37, Iss. 1, [Feb 2007]. **Suicide Terrorists: Are They Suicidal?.** Retrieved from:
    https://search-proquest-com.ezproxy.lib.ryerson.ca/docview/224869508?pq-origsite=summon

4. Keene, Shima D.**Journal of Money Laundering Control; London** Vol. 14, Iss. 4, (2011): **Terrorism and the internet: a double-edged sword.** Retrieved from:
    https://search-proquest-com.ezproxy.lib.ryerson.ca/docview/900083145?pq-origsite=summon

5. Bussel, Breanna Renee.William James College, ProQuest Dissertations Publishing, [2017]. **PATHWAY TO JIHADIST TERRORISM: A CRITICAL LITERATURE REVIEW.** Retrieved from:
    https://search-proquest-com.ezproxy.lib.ryerson.ca/docview/1966517098?pq-origsite=summon

6. Ismail Onat 1 & Zakir Gul [Dec 2018]. **Terrorism Risk Forecasting by Ideology**. Retrieved from:
    https://link-springer-com.ezproxy.lib.ryerson.ca/article/10.1007%2Fs10610-017-9368-8

7. Nilay Saiya, Anthony Scime [**Nov 2015]. Explaining religious terrorism: A data-mined analysis.** Retrieved from:
    https://journals-scholarsportal-info.ezproxy.lib.ryerson.ca/details/07388942/v32i0005/487_ertada.xml

8. Todd Sandler [March 2014]. **The analytical study of terrorism: Taking stock.** Retrieved from:
    https://journals-scholarsportal-info.ezproxy.lib.ryerson.ca/details/00223433/v51i0002/257_tasotts.xml

9. Daniel Meierrieks & Thomas Gries [January 2013]. Causality between terrorism and economic growth. Retrieved from: https://journals-scholarsportal-info.ezproxy.lib.ryerson.ca/details/00223433/v50i0001/91_cbtaeg.xml

10. Kenneth David Strang [May 2015]. **Exploring the relationship between global terrorist ideology and attack methodology**. Retrieved from:
    https://www-jstor-org.ezproxy.lib.ryerson.ca/stable/43695457?pq-origsite=summon&seq=1#metadata_info_tab_contents

11. Alberto Abadie [May 2006]. **Poverty, Political Freedom, and the Roots of Terrorism.** Retrieved from:
    https://www-jstor-org.ezproxy.lib.ryerson.ca/stable/30034613?pq-origsite=summon&seq=1#metadata_info_tab_contents

12. Andrew Barra and Alexandra Herfroy-Mischlerb [March 2017]. ISIL's Execution Videos: Audience Segmentation and Terrorist Communication in the Digital Age. Retrieved from: https://www-tandfonline-com.ezproxy.lib.ryerson.ca/doi/abs/10.1080/1057610X.2017.1361282

13. Iwa Salami [May 2017]. **Terrorism Financing with Virtual Currencies: Can Regulatory Technology Solutions Combat This?** Retrieved from: https://www-tandfonline-com.ezproxy.lib.ryerson.ca/doi/abs/10.1080/1057610X.2017.1365464

14. Matthew Simpson [July 2014]. **Terrorism and Corruption.** Retrieved from: https://journals-scholarsportal-info.ezproxy.lib.ryerson.ca/details/00207659/v44i0002/87_tac.xml

15. Todd Sandler; Walter Enders [2007]. **Applying Analytical Methods to Study Terrorism.** Retrieved from: https://onlinelibrary-wiley-com.ezproxy.lib.ryerson.ca/doi/abs/10.1111/j.1528-3585.2007.00290

16. Samuel J. Leistedt [July 2016]. **PSYCHIATRY AND BEHAVIORAL SCIENCES.** Retrieved from: https://onlinelibrary-wiley-com.ezproxy.lib.ryerson.ca/doi/full/10.1111/1556-4029.13170