Global Terrorism Analysis

Samiksha Bandbuche, Saurabh Deshpande

Madhur Jagdale, Prince Jain

Cohort Oslo Students,

Almabetter, Bangalore

Abstract:

Terrorist attacks pose a great threat to global security, and their analysis and prediction are imperative. Considering the high frequency of terrorist attacks and the inherent difficulty in finding related terrorist organizations, types of attacks and weapons used, we propose a classification framework based on ensemble learning for classifying and analyzing damages of property and life. The framework includes data collection and understanding, data cleaning and manipulation, data preprocessing exploratory data analysis (EDA). Based on a quantitative statistical analysis of terrorist organization activities in GTD from 1970 to 2017, we constructed some classification models. The proposed classifier framework is useful for analyzing types of attacks, weapons used and their impact with effectiveness which can be used to devise defense strategies to counter terrorism by defining safety protocols, equipping with necessary tools, technologies and trainings to security personnel's.

1. Problem Statement:

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. Data provided here is big source of record of continuous terrorist activities. Here we have to provide a helpful data analysis for having multilateral initiatives in response to terrorism. The following points are some major questions which we need to answer:

- Global Analysis
- Country wise Analysis
- Region wise Analysis
- Types of attacks & Weapons used

2. Analysis of Global Terrorism Dataset:

The dataset was derived from data on terrorist attacks from 1970 to 2017 from the GTD, which is managed by the National Consortium for the Study of Terrorism and Responses to Terrorism (START). The GTD dataset is considered to be the most comprehensive database for recording global terrorist activity. The information of terrorist organizations in GTD is represented by the "gname," "gname2," and "gname3" attribute fields, which, respectively, represent up to three organizations participating in an event. Most events have only a gname field value, and some events may only be represented by an unknown. Therefore, this article is focused on analyzing and predicting the "gname" attribute field (that is, the major organization). For a very small number of events with more than one terrorist organization, we focus on the major terrorist organization. According to the analysis of the dataset, there were 3,537 non repeated statistics on the attribute fields of the terrorist organizations recorded. Except for the records of unknown terrorist groups as " gname = Unknown," there were 3,536 terrorist groups in the data.

Our preliminary analysis shows that there were 181,691 identified terrorist incidents in the GTD dataset from 1970 to 2017, excluding incidents with unknown terrorist organizations. Among the cases where terrorist groups have been identified, some terrorist organizations were very active and launched numerous attacks; 19 large terrorist organizations that launched more than 1,000 terrorist attacks, 32 terrorist organizations exceeded 500, and 122 terrorist organizations exceeded 100. These 122 terrorist organizations launched 78,107 terrorist attacks, accounting for more than 79% of all known terrorist group incidents.

3. Research Methods:

In this paper, we propose a classification framework based on ensemble learning to classify and predict terrorist organizations. The framework involved four steps, including data collection and understanding, data preprocessing, data cleaning and Exploratory Data Analysis (EDA).

3.1. Data collection and understanding: The data primarily contained the following attributes of information: GTD serial number, date, event description information, time, location, attack description information, weapon information, target information, victim information, casualty information, and action results. There were many fields under each type of information to enrich the data. Each terrorist attack was stored as a record (i.e., a row) of 135 attributes such as country, year, number of deaths and injuries, and use of weapons. Among them, there were 46 attributes with a completeness of more than 70%.

3.2. Data Preprocessing: In the dataset, the average number of attacks by all terrorist organizations was 28. However, 3,230 terrorist organizations (more than 91% of all terrorist organizations) launched fewer than 28 terrorist attacks, and 2,600 terrorist organizations (73% of all organizations) launched fewer than five terrorist attacks. These 2,600 terrorist organizations launched 4,038 terrorist attacks, which accounted for only 4% of the identified terrorist attacks (i.e., attacks by identified terrorist organizations). If all terrorist organizations were predicted, too many categories and low sample categories may cause unfavorable training interference noise. Therefore, to make the experiment closer to reality and the trained model more effective, samples with fewer than five terrorist attacks were removed in this study. Some attributes are unrelated to the prediction of terrorist organizations. Training on these attributes would not only increase the required training time but also render the training results unreasonable or impractical; therefore, data preprocessing operations are essential. At this stage, the GTD dataset was processed through data cleaning, feature engineering, and data normalization.

3.3. Data cleaning: Data cleaning aims to reduce the dimensions of the GTD dataset by detecting and deleting irrelevant or redundant attributes and case records. First, attribute fields that contained descriptive text or too many missing values (the missing threshold was set to 30%) were removed. Second, missing values in specific attribute fields were filled with the numerical value corresponding to “unknown” according to the data description rules provided by the GTD. Third, some attribute fields were converted into numerical values to facilitate later processing. For example, the “related” attribute field provides the “eventid” of other terrorist attacks’ related to this terrorist attack in text format, and we convert it to the count of related terrorist attacks. The number of event records after these three steps was reduced to 98,909. Fourth, after deleting the records of terrorist attacks with fewer than five terrorist attacks, we filtered the remaining records of terrorist attacks according to five conditions (i.e., ≥5 times, ≥50 times, ≥100 times, ≥500 times, ≥1000 times). Eventually, the number of records in the experimental dataset was reduced to 94871 after the data cleaning process.

3.4. Exploratory Data Analysis (EDA): Exploratory data analysis (EDA) is used by data scientists to analyze and investigate data sets and summarize their main characteristics, often employing data visualization methods.

4. Datasets:

The primary dataset used in this project is called START (Study Terrorism and Response to Terrorism) dataset which is a part of the Global Terrorism Database, compiled by the START consortium. START dataset is an unclassified, open-sourced, freely available dataset for anyone to use and has the most comprehensive collection of terrorist events among all other available datasets. START dataset contains data of more than 180000 terrorist events happened since 1970 and has over 120 variables describing each attack. Some of the key attributes consisting those variables which are used for this project are listed below:

Attributes	Description
Region	Name of the region where the attack happened. Region consists values like East Asia, South Asia, Western Europe, etc
Country	Name of Country where the attack happened. Country consists values like Iraq, India, Pakistan, etc
Success	'1' if attack was a success. '0' if attack was a failure.
Type_Of_Attack	The type of attack happened. Type_Of_Attack consists of categories like explosion, armed assault, assassination, kidnapping, unarmed assaults.
Damage_To_Prop	Total property damage happened in any event
Killed	Number of people killed in any event
Wounded	Number of people wounded in any event

5. Analysis:

This section consists of details regarding the visual results:

5.1 Terrorist Attack World Map 1970 to 2017:

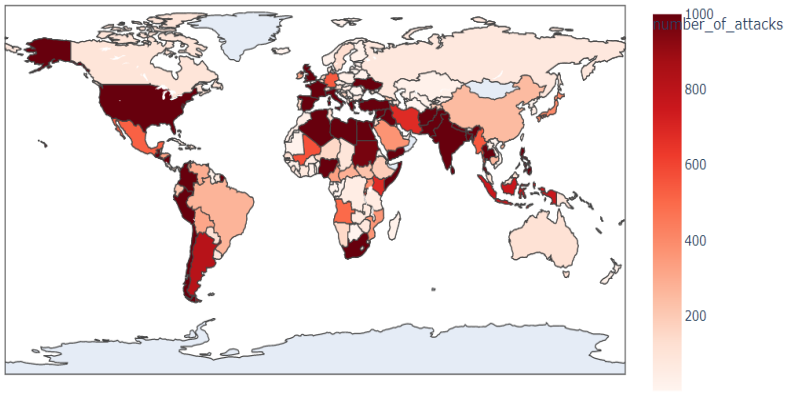

Figure 1: Terrorist Attack world map from 1970 to 2017

Here global terrorist attacks are shown with respect to time ranging from 1970 to 2017. Animation in the tool is a graphical time-lapse of all terrorist attacks from 1970 to 2017 and represents all attacks that happened in each specific year at the available geographic location in the world with a little red mark. Animation of terrorist activities helps understand the transition and spread of such activities over the years.

3.2 Terrorist Activities by Region:

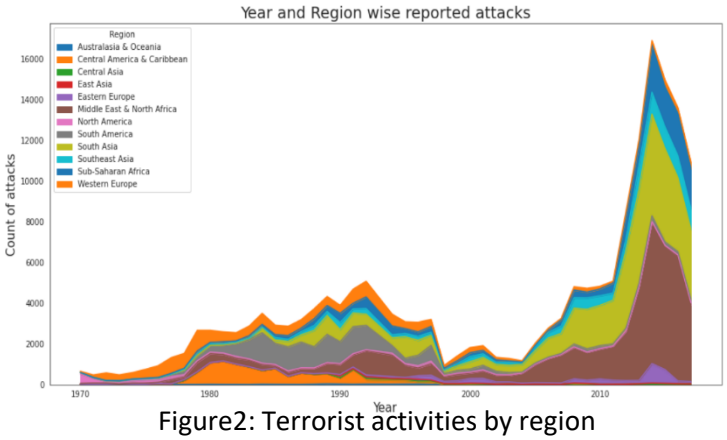

Figure2: Terrorist activities by region

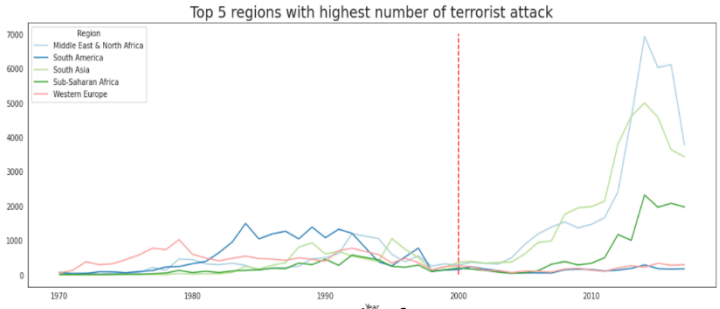

Figure3: Line Graph of activities in top 5 region

Based on the geographic location of countries, they have been subcategorized into twelve regions to compare the rate of terrorism in each one of them as shown in Figure 2 &Figure 3. From 1970s to early 2000s South America and Western Europe were on higher for reported attacks however, after 2000s we see these 2 countries with very low reported attacks. Post 2000s till 2017 we see tremendous increase in terrorist activities in 'Middle east & North Africa', 'South Asia' and ' Sub-Saharan Africa'. Terrorism here does not show an equal distribution among all regions.

3.3 Terrorist Activities by Country:

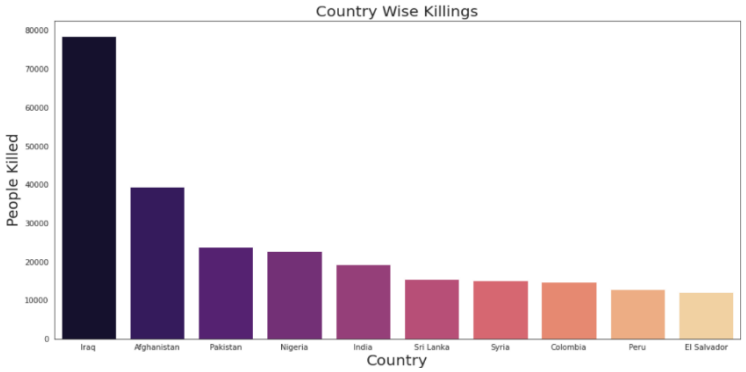

Figure4: Terrorist activities in top 10 country

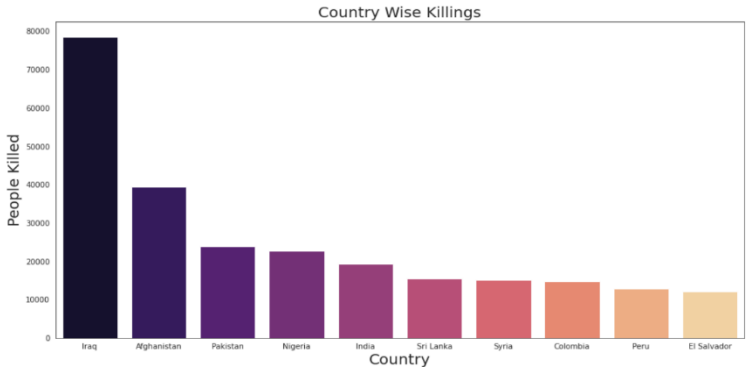

Figure5: Top 10 country with highest number of reported killings

Figure 4 shows some of the most affected countries are Iraq, Pakistan, Afghanistan, and India based on the total number of attacks. Figure 5 shows top 10 country with highest number of reported killing. Philippines, United Kingdom, Turkey these are 4 countries though in top 10 attacked countries, these countries are not in top 10 count of killings. Nigeria though not in top 10 attacked countries, it is 4th highest country with reported killings. Graph does explain how some countries are prone to violent actions and difference in an ideology which can lead to extreme terrorism.

3.4 Attacking Methods by Terrorists:

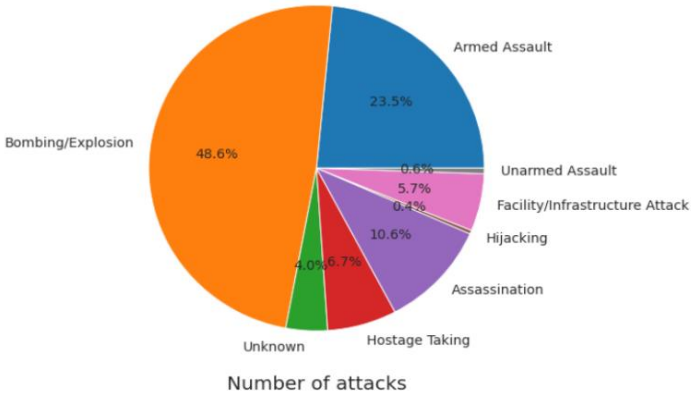

Figure6: Attacking Method by Terrorist

Different types of weapons and methods have been used by attackers. The type of attacks are unarmed assault, Infrastructure attack, kidnapping, barricade incident, hijacking, bombing/Explosion, armed assault assassination, etc. These attributes can explain which are the most often used means of attack. From above graph it very evident almost half of attacks are Bombing and explosion types, followed by armed assault and assassinations in 2nd and 3rd place respectively.

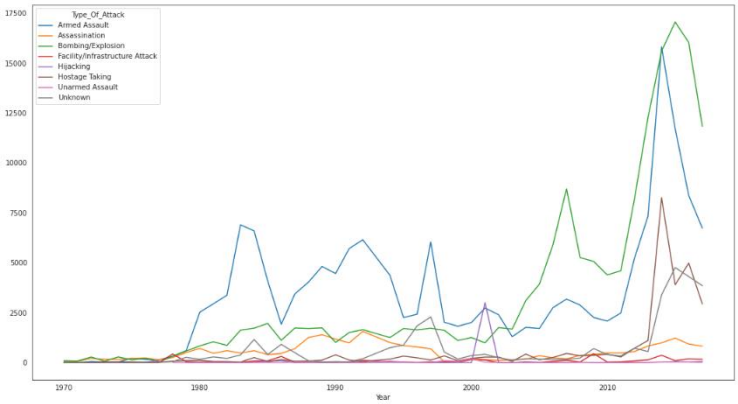

Figure7: Killings by various type of attack

From figure 7 we can see that, before 2000-2002 Armed assault was prominent reason for death of people, 2003 onwards Bombing/Explosion became the top reason for death. During 2000 to 2002 death due to hijacking were at peak.

3.5 Type of Targets:

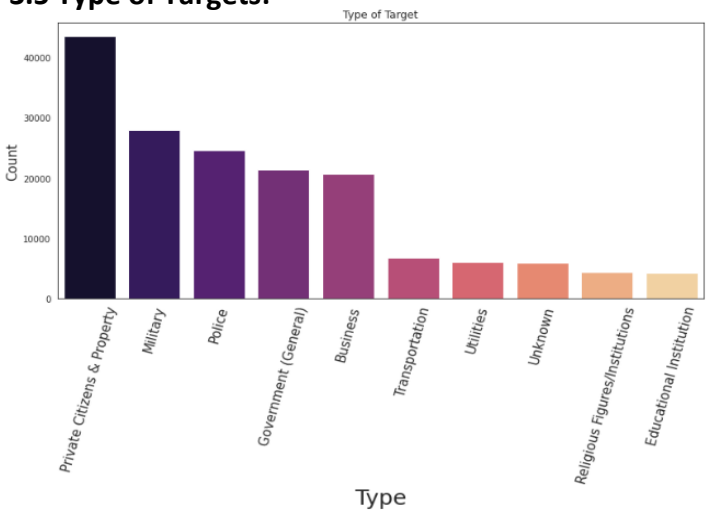

Figure7: Type of Target

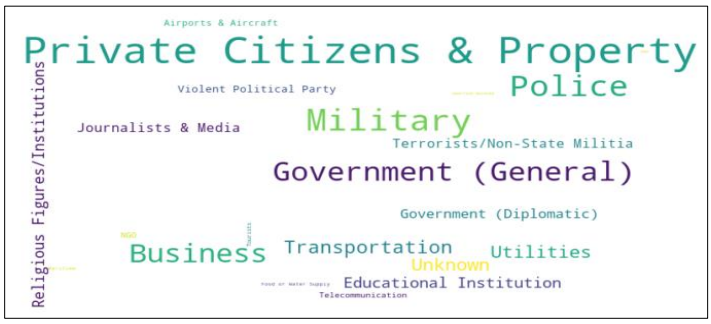

Figure8: Wordcloud for type of target

The attacker always tries to make an impact by targeting their victims. Analyzing the type of target will help understand their objective and most likely their motives. Terrorism is driven by an ideology that tries to make a change or impose an ideology. Looking into the most commonly targeted attributes will signify the attacker's objectives and terrorism in general. There are more than 100 distinct target type. These target types are generalized into 22 categorizes. Here Figure 8 shows that citizens, military, government, and police are the most common targets. This graph explains that terrorist groups or individuals have a dislike towards the authority of the state or the nation. Above plot we can see Private Citizens & properties, Military and Police are the top 3 favorite targets by terrorists.

3.6 Terrorist activities each year:

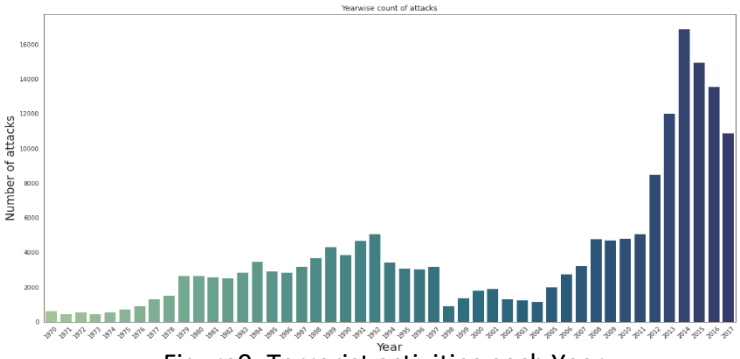

Figure9: Terrorist activities each Year

Summarizing all the terrorist attacks over the years can provide us an idea about how terrorism has evolved and what rate has it impacted the world each year? Figure 8 shows data from 1970 to 2017 for the total number of attacks happened each year. Terrorist attacks were quite low in numbers in the decade of 1970. Terrorism then had a fairly rise in the 1980s and early 1990s and was considerably low in the next decade but then terrorism rose from early the 2000s topping the charts like never before in the history. Hostile environment and global tension have increased because of the number of attacks in recent years. This observation can help investigate factors that adversely impacted the sudden rise in number of attacks. It is evident from plot 2014 has been the worst affected year in terms of number of attacks and after that we are seeing gradual decline in coming years. From year 1997 to 1998 number of attacks reduced by almost 1/3rd and attack counts were on nearly flat trend till 2004. 2004 onwards we see significant increase in terrorist activities.

4. Technologies used:

Python: Python is a high-level interpreted language that supports different platforms like Windows, Linux, Mac, Raspberry Pi, etc. Python can be used for creating web applications, database systems, handle big data, perform complex mathematical calculations. Python can be treated in an object-oriented, functional or procedural way.

Google Colab: Colaboratory, or "Colab" for short, is a product from Google Research. Colab allows anybody to write and execute arbitrary python code through the browser, and is especially well suited to machine learning, data analysis and education. More technically, Colab is a hosted Jupyter notebook service that requires no setup to use, while providing access free of charge to computing resources including GPUs.

Python packages: Following are some of the python packages used in this project.

Mathplotlib: Matplotlib is an visualiza- tion library in Python for 2D plots of arrays. Matplotlib is a multi-platform data visualiz- ation library built on NumPy arrays and designed to work with the broader SciPy stack. It was introduced by John Hunter in the year 2002.One of the greatest benefits of visualization is that it allows us visual access to huge amounts of data in easily digestible visuals. Matplotlib consists of several plots like line, bar, scatter, histogram etc.

Pandas: Pandas is an open-source library that is made mainly for working with relational or labeled data both easily and intuitively. It provides various data structures and operations for manipulating numerical data and time series. This library is built on top of the NumPy library. Pandas is fast and it has high performance & productivity for users.

NumPy: It provides structures for multiple dimensional array objects and tools for related operations. NumPy is usually used for high performance scientific computational tasks.

Seaborn: Seaborn is an visualization library for statistical graphics plotting in Python. It provides default styles and color palettes to make statistical plots more attractive. It is built on the top of matplotlib library and also closely integrated to the data structures from pandas. Seaborn aims to make visualization the central

part of exploring and understanding data. It provides dataset-oriented APIs, so that we can switch between different visual representations for same variables for better understanding of dataset**.**

Wordcloud: Word Cloud is a data visualization technique used for representing text data in which the size of each word indicates its frequency or importance. Significant textual data points can be highlighted using a word cloud. Word clouds are widely used for analyzing data from social network websites. For generating word cloud in Python, modules needed are – matplotlib, pandas and wordcloud.

5. Conclusion:

- Middle East & North Africa is the most attacked region contributing **27.78%** of reported events and **33.42%** of reported killings.
- From 'Middle east & North Africa' as well as world 'Iraq' is most attacked country contributing **13.55%** of reported events and **19.08%** of reported killings globally. 'Iraq' alone contributes to **48.8%** events and **57.1%** killings reported out of **23** countries in 'Middle east & North Africa'.
- Out of **205** countries top **10** countries have reported 61.7% of total killings. Philippines, United Kingdom, Turkey these are 3 countries though in top 10 attacked countries, these countries are not in top 10 list of killings. Nigeria, Sri Lanka, Syria though not in top 10 attacked countries, these are 4th, 6th, and 7th highest countries with reported killings.
- Private Citizens & properties, Military and Police are the top 3 favorite targets by terrorists.
- Apart from **3536** known terrorist organizations more than **27%** killings are from Unknown terrorist organizations / individuals which is big concern.
- After the Unknown type of attacker's Islamic state of Iraq and the Levant (ISIL) and Taliban are the terrorist organizations responsible for highest number of killings, both combinely contribute to more than **21%** of total killings.
- **Summary of destructions caused by Bombing/Explosion**
  - 48.6% of total events
  - 85.6% of total property damages
  - 38.2% of total killing
  - 71.1% of total wounded type casualties
  - 'Vehicle, 'Unknown Explosive Type' and 'Suicide (carried bodily by human beings)', 'Projectile (rockets, mortars, RPGs etc.)' prominent types of Bombing/Explosions
- **Summary of destructions caused by Armed Assault**
  - 23.5% of total events
  - 38.9% of total killing
  - 14.8% of total wounded type casualties
  - 'Unknown Gun Type' and 'Automatic or Semi-Automatic Rifle' prominent types of weapons used for Armed Assaults

References:

1. GeeksforGeeks
2. Analytics Vidhya
3. https://www.start.umd.edu/gtd/downloads/Codebook.pdf