Understanding Patterns of Terrorism in India Using AI Machine Learning: 2007-2017

Scott Sigmund Gartner
Pennsylvania State University (gartner@psu.edu)
Diane Felmlee
Pennsylvania State University
Rithvik Yarlagadda
University of Maryland
Dinesh Verma
IBM T. J. Watson Research Center

Overview: Terrorism represents an undesirable but seemingly inevitable part of the modern social landscape and understanding terrorism dynamics can provide useful insights for developing governance structures and policies that are both more effective at reducing violence and less invasive on general society. With the tremendous increases that are happening in Artificial Intelligence capabilities in computing technology, application of AI technologies to terrorist data can yield useful insights regarding the interaction of terrorists, governance, and society. Generally, there have been few applications of machine learning techniques to understanding patterns of terrorist behavior. Specifically, little work has been done to use AI to analyze terrorism patterns in India, which experiences among the world's highest levels of terrorism. Using the Global Terrorism Database and the South Asian Terrorism Portal we apply "shallow machine learning models" that require only a modest amount of data to train themselves and can facilitate our exploration of three questions crucial to understanding the complex dynamics of terrorism, state and society: From a description of the attack can we identify the likely terrorist group? Can we predict the likely location for the next attack from a history of past attacks? Can we identify the principal factors that cause a city to be targeted? We believe that this project will: provide an example of socially-relevant AI research; expand our understanding of the factors that shape counterterrorism policy and contribute to our greater recognition of the interwoven relationship of technology, knowledge, and society.

Terrorist incidents are an undesirable but inevitable part of the modern social landscape, and understanding the patterns of terrorist incidents can provide useful insights on mechanisms to prevent them.[1] Conflict assessment, however, has long represented both a policy and critically an analytical challenge.[2] With the tremendous increases that are happening in Artificial Intelligence capabilities in modern computing technology, application of AI technologies to terrorist data can yield useful insights to understand their patterns.[3] These approaches might allow us to identify the answers of questions crucial to understanding the dynamics of terrorism and the requirements for effective counterterrorism, such as:

- From the description of the attack, can we determine the identity of the likely terrorist group?
- Can we predict the likely location for next attack from a history of past attacks?
- Can we identify the principal factors which cause a city to be targeted?

Although mathematical techniques have been applied to understand terrorist data in various countries, there has not been a significant application of machine learning techniques to understand patterns of terrorist behavior, and specifically, little work has been done to understand the terrorism pattern in India sufficient to address these types of questions.

In past work, the authors have applied 2-mode network analysis to understand the behavior of Islamic terrorist groups in India.[4] Such an analysis has provided useful insights into the groups that are most active and the clustering in the location of the attacks that such groups undertake. However, several interesting questions remain unanswered, such as the ability to predict the potential responsible group from a description of a terrorist incident, the ability to predict the primary factors that could lead to a city becoming the target of an attack, or the ability to predict the location or time of a potential future attack. Our objective is to draw on our extant co-authored research on terrorist implicit networks[3], and author's work on AI,[5] Social Network Analysis,[6] and Terrorist Studies,[1] to build an AI Model that will address these questions.

Machine learning techniques provide this capability of building models with predictive behavior, and an interesting question is to understand the capabilities in prediction that can be obtained that can be obtained by applying machine learning to conflict events. In this paper, we would report on the insights gained from the applications of machine learning techniques to terrorist incidents in India.

The primary sources for data are: Global Terrorism Database (GTD), which is an open-source database maintained by the National Consortium for the Study of Terrorism and Responses to Terrorism (START); and South Asian Terrorism Portal (SATP). One of the challenges in applying machine learning to

---

[1] Scott Sigmund Gartner. August 2015. "Net Assessment 2.0: The Net Assessment of Violent Non-State Actors." *CTX: Conflict Terrorism Exchange.*

[2] Scott Sigmund Gartner and Catherine Langlois. "Unbalanced Policy Priorities and the Interrogation of Terror Suspects." *Foreign Policy Analysis*, Volume 14, Issue 1, 1 January 2018, Pages 107–126; Scott Sigmund Gartner. 1999. *Strategic Assessment in War*. New Haven: Yale University Press.

[3] Scott Sigmund Gartner. "Big Data Could Uncover Clue on Marathon." *USA Today*. Print: 8:A April 17, 2013 (Online 4/16/13).

[4] Rithvik Yarlagadda, Diane H. Felmlee, Dinesh Verma, Scott Sigmund Gartner. (2018) Implicit Terrorist Networks: A Two-Mode Social Network Analysis of Terrorism in India. In: Thomson R., Dancy C., Hyder A., Bisgin H. (eds) Social, Cultural, and Behavioral Modeling. SBP-BRiMS 2018. Lecture Notes in Computer Science, vol 10899. Springer

[5] Dinesh Verma, Greg Cirincione, Tien Pham, Bong Jun Ko, "Generation and management of training data for AI-based algorithms targeted at coalition operations," Proc. SPIE 10635, Ground/Air Multisensor Interoperability, Integration, and Networking for Persistent ISR IX, 106350U (4 May 2018).

[6] Diane H. Felmlee and Robert Faris. August 20, 2016. "Toxic Ties: Networks of Friendship, Dating, and Cyber Victimization." *Social Psychology Quarterly*. Vol 79, Issue 3, pp. 243 – 262.

terrorism data is that the data set is usually too small to train deep learning models such as various neural network variants. Furthermore, some of the attributes that may be drivers of terrorist incidents may not be available in terrorism data such as the ones available from the Global Terrorism Database (GTD). In order to alleviate this challenge, we have opted to apply shallow machine learning models such as decision trees and random forest to build models that require only a modest amount of data to train themselves. Furthermore, we have augmented the data available from the GTD with additional attributes of locations and perpetrator groups that may help us obtain better insights into the group behavior.

We believe that our efforts in this project will advance our understanding of technology, Knowledge, and society generally, and provide a specific example of research that can contribute to our understanding of the social impact of AI and how it affects policies and governance that impact social change.

India

India regularly experiences globally high rates of terrorist activity, ranking 7th highest in the world for terrorist incidents, a level largely exceeded mainly by war-strewn countries such as Syria, Nigeria, and Iraq.[7] Despite the high level of activity, terrorist activity within India has only been sparingly studied, with social network approaches especially rare (for an exception, see Basu 2005; Saxena et al. 2004). In recent years, India has witnessed a phenomenal rise in both the global economic and political arenas, thus, heralding the entry of this vast and highly diverse country into the league of international powerhouses. Religious diversity is one of the most important defining characteristics of India. According to the 2011 Census of India, nearly 80 percent of the population practices Hinduism, the majority religious group in the country. This is followed by Islam which comprises almost 14.2 percent of the entire population.[8] India has the world's second-largest Muslim population, although they are part of a minority group within most of the country.[9] Jammu and Kashmir (J&K) is the only state within India where Muslims are in the majority.[10] Ironically, this unusual religious make-up of the state coupled with high unemployment and government insensitivity were among the root causes that led to severe unrest in the J&K region over many decades.[11] The Kashmir conflict existed since 1947, when both India and Pakistan obtained independence from the British. Ever since, this region has been the bone of contention between the two countries which also resulting in three major wars fought. India has also recently a strong rise in Marxist violence, which was focused on North East and Deccan Plateau.

Literature Review

Terrorism in India

Terrorism is one of the most important security challenges concerning India. Multiple large-scale insurgencies mostly operated by Islamist militants, ethnic-based separatist groups, and Naxalites have made India a terrorist hot spot for over two decades now. Few studies have used tools such as social network analysis (SNA) to map out the violence patterns of different terrorist groups within India.[12]

---

[7]Global Terrorism Database. University of Maryland, July 2018, www.start.umd.edu/gtd/downloads/Codebook.pdf.

[8]"Religion Census 2011." Religion Data - Population of Census 2011 India, Census Organization of India, 2011, www.census2011.co.in/religion.php.

[9] Desilver, Drew., and David Masci. 2017. "World's Muslim population more widespread than you might think." *Pew Research* Center. Accessed from http://www.pewresearch.org/fact-tank/2017/01/31/worlds-m uslim-population-more-widespread-than-you-might-think/

[10] "Religion Census 2011." Religion Data - Population of Census 2011 India, Census Organization of India, 2011, www.census2011.co.in/religion.php.

[11] "Kashmir: Why India and Pakistan fight over it." *BBC News*. Accessed from http://www.bbc.com/news/10537286

[12] Rithvik Yarlagadda, Diane H. Felmlee, Dinesh Verma, Scott Sigmund Gartner. (2018) Implicit Terrorist Networks: A Two-Mode Social Network Analysis of Terrorism in India. In: Thomson R., Dancy C., Hyder A., Bisgin H. (eds) Social, Cultural, and Behavioral Modeling. SBP-BRiMS 2018. Lecture Notes in Computer Science, vol 10899. Springer, Cham

While most of these studies identify the patterns of violence, they fall short in analyzing the various factors that lead to terrorism in India. One of the dominant factors that contributed to the terrorist violence in India is the presence of unresolved conflicts mainly in the Jammu and Kashmir (J&K) and North-East regions. As studies show, states with poorly addressed political disputes experience a high degree of terrorism.[13] While this resonates with the macro-level explanations of terrorism, it does not really explain the variation within India. Scholars have increasingly started using more fine-grained data to examine the sub-national patterns of terrorism. For instance, one recent study argues that state-level differences in political party systems can potentially explain the variation in terrorist activity across different states within India. In particular, they show that states with stable two-party systems and majority party rule are less likely to experience terrorism than those with more fractionalized electoral systems and minority governments at the helm.[14]

General Themes in the Study of Terrorism

GTD defines terrorism as an act of violence carried out by non-state actors to mainly attain political, economic, religious, or social goals through fear or coercion.[15] Terrorism can either be transnational or domestic in nature. However, GTD's definition does not really tease out these differences, which can be problematic especially for the quantitative study of terrorism. Few recent studies identify this problem and suggest ways that would enable us to separate the GTD terrorism data into international and domestic types.[16] There is extensive research that examines the individual factors influencing each of the two different forms of terrorism. Some of the primary determinants predicting the likelihood of transnational terrorism across states include democracy,[17] ethno-religious diversity[18], economic globalization[19], and experience state failure[20]. On the contrary, some of the primary factors that influence domestic terrorism are education,[21] and minority economic discrimination.[22]

The occurrence of terrorist activities in some locations but not in others, clearly demonstrates that terrorism is a geographically driven phenomenon. This spatial variation leads to different hot spots of terrorism both transnationally[23] and within states. In analyzing the variation of terrorism over space, few studies indicate how most terrorist groups target locations that are more accessible from their own home

---

Basu, A. (2005, June). Social network analysis of terrorist organizations in India. In *North American Association for Computational Social and Organizational Science (NAACSOS) Conference* (pp. 26-28). NAACSOS.

Saxena, S., Santhanam, K., & Basu, A. (2004). Application of social network analysis (SNA) to terrorist networks in Jammu & Kashmir. *Strategic Analysis*, *28*(1), 84-101.

[13] Piazza, J. A. (2009). Economic development, poorly managed political conflict and terrorism in India. *Studies in Conflict & Terrorism*, *32*(5), 406-419.

[14] Piazza, J. A. (2010). Terrorism and party systems in the states of India. *Security Studies*, *19*(1), 99-123.

[15] p. 10, GTD Codebook. Accessed at https://www.start.umd.edu/gtd/downloads/Codebook.pdf

[16] Enders, W., Sandler, T., & Gaibulloev, K. (2011). Domestic versus transnational terrorism: Data, decomposition, and dynamics. Journal of Peace Research, 48(3), 319-337.

[17] Li, Q. (2005). Does democracy promote or reduce transnational terrorist incidents?. *Journal of Conflict resolution*, *49*(2), 278-297.

[18] Piazza, J. A. (2006). Rooted in poverty?: Terrorism, poor economic development, and social cleavages. *Terrorism and political Violence*, *18*(1), 159-177.

[19] Li, Q., & Schaub, D. (2004). Economic globalization and transnational terrorism: A pooled time-series analysis. *Journal of Conflict Resolution*, *48*(2), 230-258.

[20] Piazza, J. A. (2008). Incubators of terror: Do failed and failing states promote transnational terrorism?. *International Studies Quarterly*, *52*(3), 469-488.

[21] Brockhoff, S., Krieger, T., & Meierrieks, D. (2015). Great expectations and hard times: The (nontrivial) impact of education on domestic terrorism. *Journal of Conflict Resolution*, *59*(7), 1186-1215.

[22] Piazza, J. A. (2011). Poverty, minority economic discrimination, and domestic terrorism. *Journal of Peace Research*, *48*(3), 339-353.

[23] Braithwaite, A., & Li, Q. (2007). Transnational terrorism hot spots: Identification and impact evaluation. *Conflict Management and Peace Science*, *24*(4), 281-296.

bases and international borders and are relatively closer to symbolic centers of government.[24] Studies also show that regime-type[25] and regime-age[26] are a few other important factors that influence the likelihood of terrorist attacks across states. Extending the previous argument that terrorism violence is restricted to certain specific locations (or, hot spots), some scholars argue that terrorism could also diffuse from one state to another.[27] On the contrary, a location more prone to violence does not necessarily imply that all terrorist groups target it. As a result, there is a need to assess the variation in violence patterns among different terrorist groups. Previous studies have analyzed the behavioral patterns of violence of different terrorist groups by focusing on certain important factors such as their target choice[28], and lethality or intensity of violence.[29]

Theory

Our theoretical approach has two parts. We believe that location plays a major role in determining terrorist activity. Generally, while terrorist groups may claim to act in support of abstract, ideological causes, local factors are critical for three reasons. First, terrorist groups require logistical support such as weapons, money, training, food, housing, intelligence, which are often provided by neighboring states or groups in neighboring states. Second, local grievances frequently serve as the motivations or catalysts for distinct terrorist acts. Third, per our earlier research, we determine that capitals have a symbolic importance that makes them more likely to experience terrorism.[30] When combined, these three factors provide terrorism with a distinctly local dynamic that we explore through examining geographical attributes of terrorism patterns.

A recent study shows that certain local-level factors such as the occurrence of civil violence, sub-national economic activity, and proximity to symbolic locations like capitals and urban areas are important in explaining the patterns of transnational terrorism.[31] Furthermore, they argue that the increase in likelihood of transnational terrorism due to civil conflict is more pronounced in the areas that are close to state capitals than those that are farther away. This suggests that terrorists would prefer to target specific locations holding greater political value which in turn would help them garner more global attention to their cause.

Despite the view commonly portrayed in the media, the vast majority of terrorism is domestic. Thus, understanding the composition of the population is crucial for understanding variation in terrorism activity across space and time. We examine a number of critical demographic factors such as religion, population and population growth, literacy, and ethnicity. Previous studies show that ethno-religious

---

[24] Berrebi, C., & Lakdawalla, D. (2007). How does terrorism risk vary across space and time? An analysis based on the Israeli experience. *Defence and Peace Economics*, *18*(2), 113-131.
[25] Wilson, M. C., & Piazza, J. A. (2013). Autocracies and terrorism: Conditioning effects of authoritarian regime type on terrorist attacks. *American Journal of Political Science*, *57*(4), 941-955.
[26] Piazza, J. A. (2013). Regime age and terrorism: Are new democracies prone to terrorism?. *International Interactions*, *39*(2), 246-263.
[27] Cliff, C., & First, A. (2013). Testing for contagion/diffusion of terrorism in state dyads. *Studies in Conflict & Terrorism*, *36*(4), 292-31.
[28] Asal, V. H., Rethemeyer, R. K., Anderson, I., Stein, A., Rizzo, J., & Rozea, M. (2009). The softest of targets: A study on terrorist target selection. *Journal of Applied Security Research*, *4*(3), 258-278.
[29] Asal, V., & Rethemeyer, R. K. (2008). The nature of the beast: Organizational structures and the lethality of terrorist attacks. *The Journal of Politics*, *70*(2), 437-449.
[30] Rithvik Yarlagadda, Diane H. Felmlee, Dinesh Verma, Scott Sigmund Gartner. (2018) Implicit Terrorist Networks: A Two-Mode Social Network Analysis of Terrorism in India. In: Thomson R., Dancy C., Hyder A., Bisgin H. (eds) Social, Cultural, and Behavioral Modeling. SBP-BRiMS 2018. Lecture Notes in Computer Science, vol 10899. Springer, Cham
[31] Marineau, J., Pascoe, H., Braithwaite, A., Findley, M., & Young, J. (2018). The local geography of transnational terrorism. *Conflict Management and Peace Science*, 0738894218789356.

diversity plays a big role in determining terrorism.[32] While this study empirically tests its argument on a cross-national dataset, we believe that a similar logic can be applicable to the study of transnational terrorism sub-nationally. As a result, we expect demographic factors such as the breakdown of district-level population based on religion, urban/rural, and literacy levels, to have an influence on the occurrence of terrorist incident. When combined together, we think the aforementioned factors can manifest critical, but sometimes hard to observe patterns and relationships that influence terrorism and society dynamics

Data Analysis on Indian Terrorism Data
From this report, we analyzed the terrorism incident database relevant to India to answer the following questions:

1. Can we predict the important factors that contribute to the occurrence of a terrorist incident in a district?
2. Can we create an AI model that can predict whether or not a district will have a terrorist incident?
3. Can we create an AI model that can predict the intensity of terrorism incidents in a district of India?
4. Can we create an AI model to predict the perpetrator of a terrorism incident?
5. Can we predict the important factors that can determine the identity of an organization that perpetrated a terrorist incident?

Method
Our machine learning models have been built using the open source sci-kit learn library (http://scikit-learn.org/) , and provides a way to understand the clustering of the incidents, and to predict the most likely perpetrator groups when we only have limited details available on a new occurrence. In this paper, we report on the effectiveness of the different techniques in machine learning models to be able to make such predictions accurately.

Data Overview
The data for this analysis consisted of two components, the first dealing with terrorist events, and the other dealing with the demographics of districts where the incidents happened. Information about events was extracted from the Global Terrorism Database. The district data was extracted programmatically from the 2011 census of India which was available online at https://www.census2011.co.in. District latitude and longitude information was extracted from the Wikipedia entries corresponding to them, and for a few where Wikipedia entry did not exist, determined from online mapping site of latlong.net

The district data consisted of 640 entries, one for each entry consisting of the following information:
- Whether the district consisted of a state capital?
- The total population, and breakdown of population by different religions.
- The literacy, population growth in the district, and the breakdown of rural and urban, literacy rates in each district.
- The latitude and longitude of the district.

Of the 640 districts in India, between the time-period of 2007-2017, 414 districts were subject to at least one incident of terrorist violence. These incidents include those where the identity of the perpetrator was unknown.

Note, critically, we included demographic information from all districts – whether or not they experienced a terrorist incident during our period of analysis. By including these 'non-event' data we are able to

[32] Piazza, J. A. (2006). Rooted in poverty?: Terrorism, poor economic development, and social cleavages. *Terrorism and political Violence*, *18*(1), 159-177.

develop a full and robust statistical portrait of demographic dynamics. Put simply, we create independent variables and measures that are not driven by the selection of our dependent variable. This type of approach is highly unusual for studying these types of dynamics and this study represents one of the first to employ this more comprehensive and subsequently less statistically biased data approach.

The event data was extracted from the GTD and each entry corresponded to one of the terrorism incidents in India. The following information was associated with each of the entries:

- *Incident location*: We aggregate the location of each terrorist event to three different spatial/administrative units. They include the state or province, district, and city in which an event occurred. Additionally, we also include the latitude and longitude that would enable us to analyze precisely the geographic patterns of terrorism in India. The city and state names were available from the GTD database. They were edited manually to match with the names used by the Census of India. The district information was determined by mapping the latitude and longitude of the incident to the closest district in the same state.

- *Time of attack:* To further our aim of examining the spatiotemporal patterns of terrorism, we include a variable that can capture the time of the attack. Here, we aggregate the events to yearly temporal units.

- *Capital:* This binary variable is used to indicate whether a given terrorist incident occurs in a capital city or not. It is coded as '1' if the event location is a capital city and '0' otherwise.

- *Type of attack:* This variable captures the general method of attack in a given terrorist event. According to the GTD, attack types can be categorized into nine different types. Each event is coded with up to three different attack types. While in most cases, there is only one attack type recorded for each event, there could be some events each with multiple attack types. This usually occurs when any given attack comprises a sequence of events. For events with multiple attack types, GTD follows a hierarchical ranking of the nine different attack categories and accordingly ranks the three attack types for each event.[33]
  In our analysis, we considered each attack type as a distinct binary variable that records the presence (coded as '1') or absence (coded as '0') of that attack type in a given event. The type of attack that was attempted, one of Assassination, Hijacking, Kidnapping, Barricade, Bombing/Explosion, Armed assault, Unarmed assault, Facility/Infrastructure. Furthermore, we also include a count variable to track the number of attack types coded for each event. The encoding used was identical to that of GTD.

- *Type of weapon:* This variable records the type of weapon used in a given terrorist incident. The GTD classifies weapon types into 13 different categories. Each event is coded with up to four different weapon types. While in most cases, there is only one weapon type recorded for each event, there could be some events each with multiple weapon types.[34] In our analysis, we considered each weapon type as a distinct binary variable that records the presence (coded as '1') or absence (coded as '0') of that weapon type in a given event. Additionally, we also included a count variable to track the number of weapon types that each individual event has. The weapons used in the attack, one of Biological, Chemical, Radiological, Nuclear, Firearms, Explosives, Fake Weapons, Incendiary, Melee, Vehicle, or Sabotage equipment.The encoding used was identical to that of GTD.

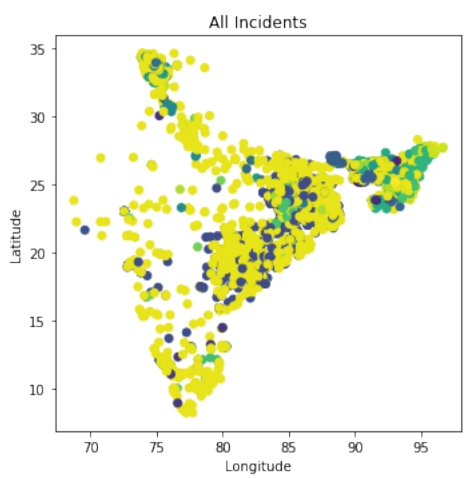
[33] p. 23, GTD Codebook. Accessed at https://www.start.umd.edu/gtd/downloads/Codebook.pdf
[34] p. 28-31, GTD Codebook. Accessed at https://www.start.umd.edu/gtd/downloads/Codebook.pdf

- *Perpetrator Group Name:* This variable contains the name of the perpetrator group that carried out the terrorist attack. Each event is coded up to three different group names to account for the terrorist incidents that involve multiple perpetrator groups claiming responsibility.

- *Intensity of attack:* We use this variable to capture the impact of the terrorist incident measured by the number of fatalities and injuries.

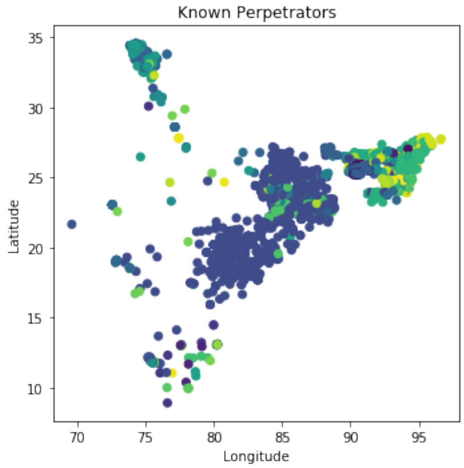
For our analysis, we aggregated event locations to their corresponding districts. This is the smallest administrative unit that can be coded with consistency across all the events.[35] As a result, we constructed a second dataset that is specific to 640 districts within India. In the district dataset, we are primarily interested in examining the role of district-level demographics in predicting the occurrence of terrorist attacks. Since the demographic data was from 2011, we examined the incidents that happened in the decade surrounding it, namely events from 2007-2017 were considered. There were 7697 such terrorist incidents, among which the identity of the perpetrating group was known in the case of 3744 incidents. From the GTD data, we mapped the identity that was expressed in general terms (e.g. as Insurgents, or Sikh extremists) without identifying the exact group (such as United Liberation Front of Assam) were considered as unknown groups.

The figure below shows the location of the different incidents in the decade, with the yellow dots showing the incidents where the perpetrator was not known. There were 3953 entries in which the perpetrating organization was not known. One of the goals in creating the model was to have the ability to predict the perpetrator organization for the entries in which the perpetrator was unknown. From the map, the location of the unknown perpetrators is distributed throughout the length and breadth of the country. The unknown perpetrators are shown as the yellow dots.



The figure below shows the events conducted by one of the known perpetrators. The known events were mostly located towards the Eastern side of the country and in the Kashmir Valley.

[35] The administrative unit of city as coded by GTD is not consistent across all events. Some events are coded to the city level while others are coded to the district level. Thus, to maintain consistency, we used district level as our unit of analysis.



There were 166 perpetrator organizations in total among all of the incidents that happened between 2007 and 2017. The bulk of the known incidents can be attributed to the rise in the Naxalite activities in India, spearheaded by the Communist Party of India. These incidents resulted in the large cluster visible along the Eastern part of India.

Descriptive Data Results

District Analysis

The district analysis objective was to identify whether we could determine the conditions which lead to the occurrence of the incident in the district. We identified these conditions by building an AI model which could identify the factors that allowed for the differentiation between the different types of districts.

There are many different types of AI models that could have been used for the analysis. However, since we only have about 600 points to train an AI model, the most prevalent type of model, deep models such as neural networks of different nature, were not suitable. These models generally require a much larger number of points to train them. Furthermore, it is hard to explain the contribution of each input into the result of the neural network classifier.

Given these constraints, we opted to use two common models which fall into the categories of "shallow AI models". These can be trained with considerably less amount of data. Furthermore, we chose the model types that could provide an important metric for the input features influencing their output. Decision trees and random forest were the two type of models we chose based on these factors.

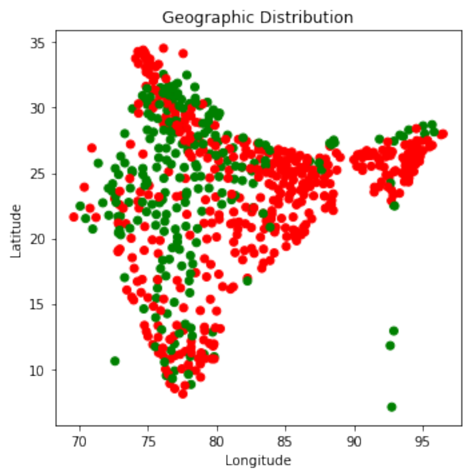
In order to create a prediction, we augmented the input available to also compute the fraction of the total population which belonged to each of the religious types. Thus, the input features that were used to create the AI models were the presence of the capital in the district (a binary value), and the demography fields (total population, populations of Hindus, Muslims, Christians, Sikhs and others, as well as the fraction of the population that were Hindus, Muslims, Christians, Sikhs and others).

We tried to predict the occurrence of terrorism incidents in two ways, in the first way we simply tried to predict whether or not a district will have a terrorist incident. In the other way, we classified districts with incidents into four categories based on the number of incidents, defined as per the following table:

Intensity Level	Definition	Corresponding Number of Districts

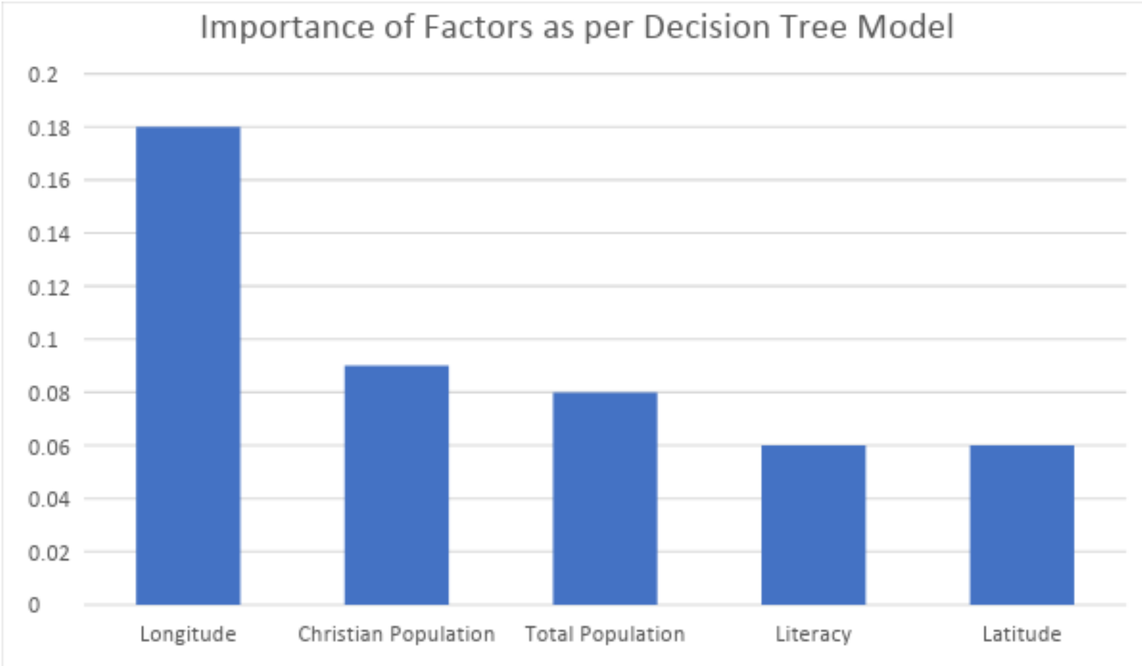
Free	No incidents in district	226
Low	Less than 5 incidents in district	215
Medium	More than 5 but less than 10 incidents	61
High	More than 10 but less than 25 incidents	60
Extreme	More than 25 incidents	78

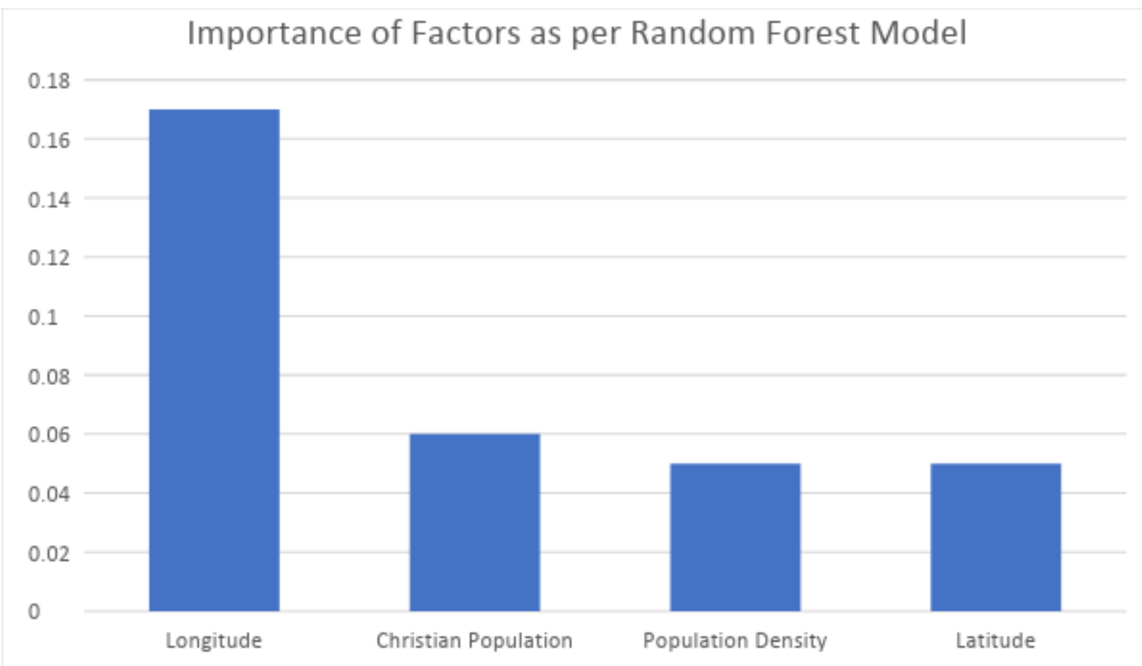
The distribution of the districts subject to terrorist incidents (red) and those free of incidents (green) are as shown below:



Factors Contributing to Incidents:

We created both a decision tree and random forest model to determine how much importance the model provided to each of the different contributing features. The importance is a measure of the usefulness of the feature to make the decision properly. We only show the top four features (or five if fourth and fifth features have the same importance). For the binary prediction whether or not the district will have an incident, resulted in the following importance from the decision tree model:



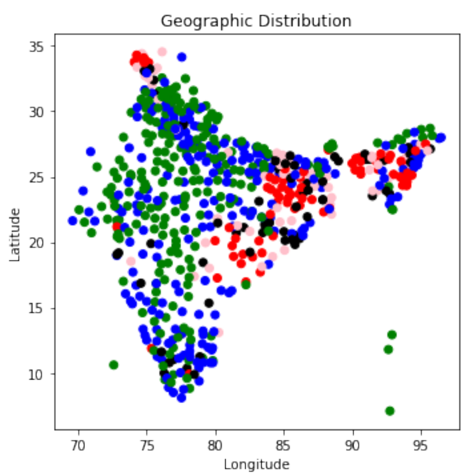


As can be seen from the importance table, the factor that stands out as most important from the point of predicting the occurrence of a terrorist incident is the location of the district, where the longitude has a more important role to play than the latitude. The rationale will be obvious from the distribution of the incidents, which are prevalent more towards the Eastern side of the country, but spread throughout from the North to the South. As a result, latitude, while important as shown in Random Forest data, is less important than longitude in providing a discrimination between districts that experienced terrorist violence and ones that did not.

One factor that was missing was whether the district contained a capital. The fact that the presence of the capital does not contribute much to the prediction of the incidence was a surprise since most of the incidents happen in capital locations. Of the 35[36] capital cities in India, 11 were subject to terrorist incidents which is a much higher percentage (31.4%) than the non-capital states (46 out of 604 or 7.6%). However, other factors had a more important role in discriminating whether or not a district will be incident free.

Total Population (and closely related population density) plays a role in determining whether or not an incident will take place, as does the total number of Christians in a district. While both of these are surprising demographic factors, the Naxalite incidents in 2007-2017 happen to be in the areas which are relatively sparsely populated, and also have a larger number of Christians, e.g. the North East, and the trial districts in the Deccan Plateau. Thus, these two factors have a strong discriminative value in determining whether or not a terrorist attack happens.

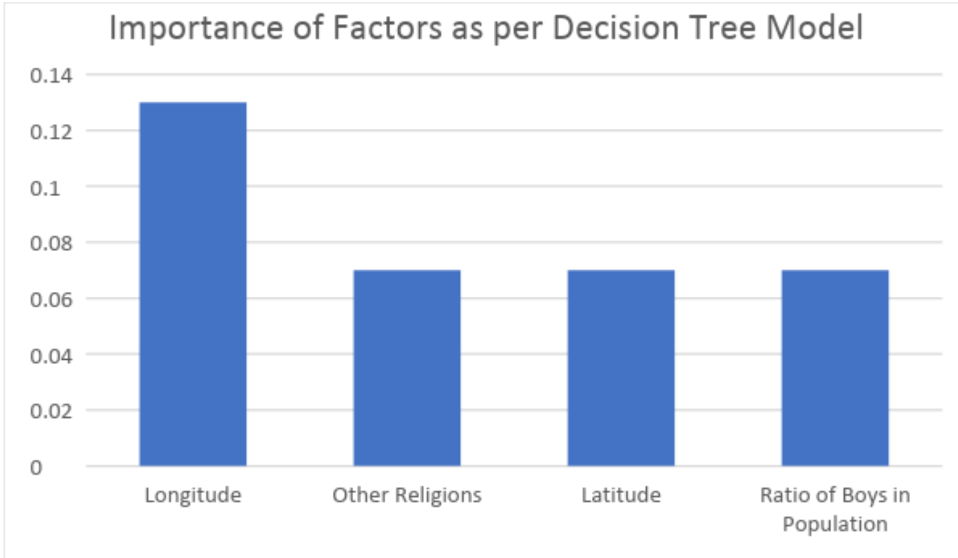
We wanted to explore whether the same factors would turn out to have the predictive value when we consider the prediction of the intensity of terrorist activity instead of a binary decision. The distribution of the attacks is as shown below:

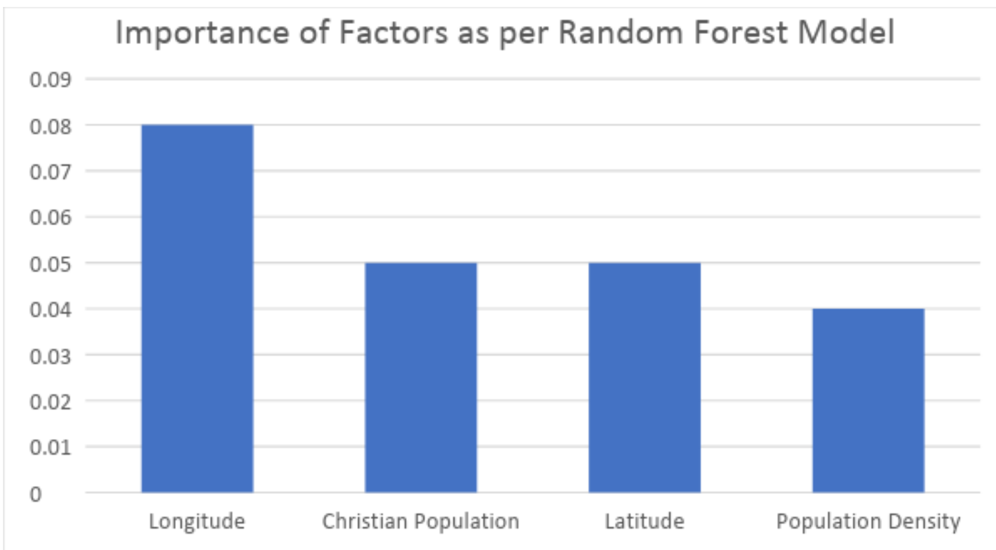


In this figure, the green dots indicate no attacks, the blue dots are districts with 5 or less attacks in the decade, the black ones are those with 6-10 attacks, the pink ones have 10-25 attacks and the red ones have 25+ attacks in the 10-year period that was studied.

When plotting the same predictive power for the decision tree and random forest models for predicting terrorist incidents, we come with the following figures:

[36] India has 29 states and 7 Union Territories (including Delhi), but two states (Punjab and Haryana) share a capital, resulting in 35 capital cities.





While the actual importance numbers are somewhat different in this set of predictions, the same relative importance statement holds true for these predictive models as well.

Looking at the four charts above, we can determine that the location of a district (its latitude and longitude) have a strong influence on whether or not it experienced a terrorist incident. Furthermore, the total population in the district and the minority religions (Christian or a non-traditional religion) in the district have a significant discriminative value in determining whether or not an incident would happen.

Accuracy of Incident Models

To understand whether the decision tree models and the random forest model that result from the data are good enough to predict the occurrence of terrorist incidents, or the intensity of such incidences, we used the models to see how well they perform in two ways, first by training the model on the full data, and seeing how well they predict the results on the same data, and secondly by training the model only on 90% of the data and checking how well they perform on the remaining 10% of the data not used during training.

In each of these approaches, the process was repeated for 100 iterations to get a sense of the variability in results that would be obtained.

To compare how well the two AI models performed, we will compare them against the baseline approaches of an Intelligent guesser. The intelligent guesser would look at the distribution of statistics for the outcome in the training set and guess the results of a testing set using that distribution. It provided a base-check against which the performance of the models can be compared. For accuracy measure, the match with the predicted and known result was used. We tried to use other measures of accuracy such as F-score, but with the small sample size for testing, conditions for calculating those scores were frequently not met. Therefore, we opted for the more basic but calculable metric of accuracy.

For the binary classification model, we get the following results over 100 iterations:

Model	Minimum Accuracy	Maximum Accuracy	Mean Accuracy	Standard Deviation
Decision Tree	100%	100%	100%	0.00
Random Forest	97%	100%	99%	0.5%
Intelligent Guesser	50%	59%	54%	2%

For the intensity classification model over five classes, we get the following results over 100 iterations:

Model	Minimum Accuracy	Maximum Accuracy	Mean Accuracy	Standard Deviation
Decision Tree	100%	100%	100%	0.00
Random Forest	97%	99%	98%	0.5%
Intelligent Guesser	23%	31%	27%	1.7%

In both of these cases, the AI models out-perform the random or the intelligent guesser. Furthermore, the variability in prediction accuracy is much lower for the AI models compared to guessing.

One may expect that an AI model tested on the same data that it is trained on to perform fairly well. Therefore, we tested it on the data by dividing it into 90% for training and 10% for testing.

For the binary classification model, we get the following results over 100 iterations:

Model	Minimum Accuracy	Maximum Accuracy	Mean Accuracy	Standard Deviation
Decision Tree	56%	69%	62%	2.4%
Random Forest	64%	84%	75%	3.6 %
Intelligent Guesser	34%	67%	54%	5.6%

For the intensity classification model over five classes, we get the following results over 100 iterations:

Model	Minimum Accuracy	Maximum Accuracy	Mean Accuracy	Standard Deviation
Decision Tree	31%	56%	44%	4.2%
Random Forest	22%	59%	42%	5.6%
Intelligent	0%	56%	27%	7.6%

Guesser				

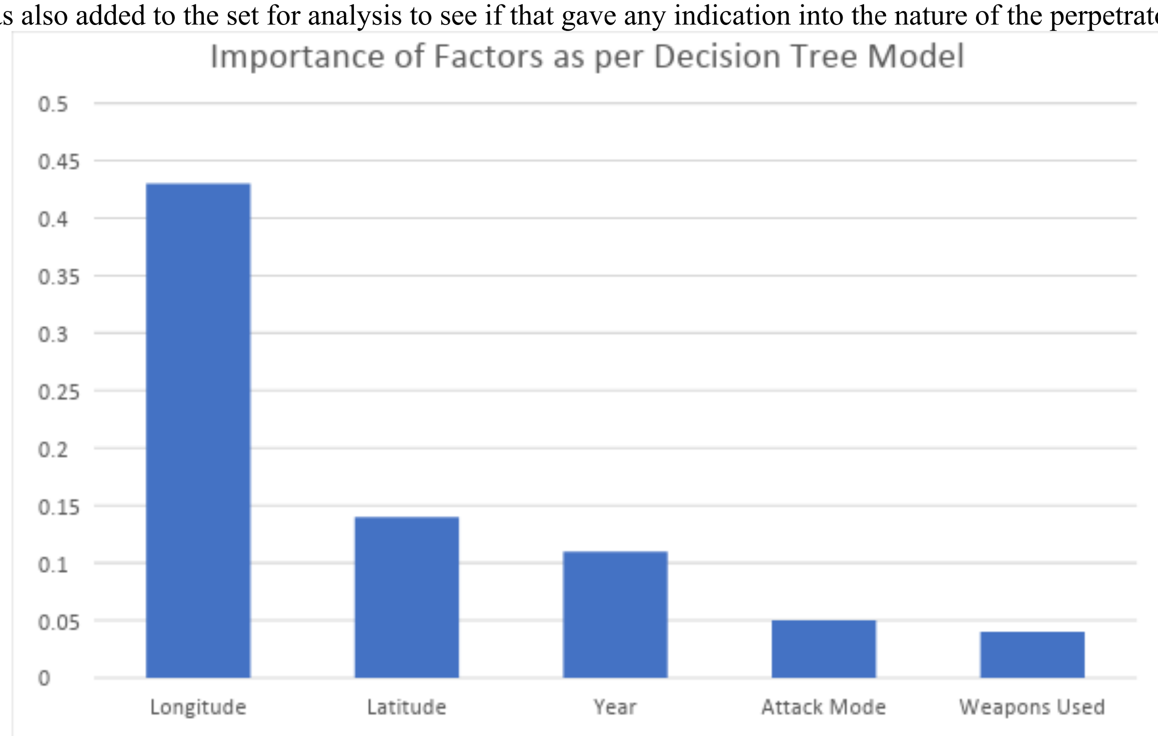
The AI model are doing considerably better than intelligent guessing.

Perpetrator Analysis

During the perpetrator analysis, we are interested in predicting the identity of the perpetrator from the data regarding each of the events. As in the case of district analysis, with about 3000 points to train an AI model, a deep AI model would not be trained well. The AI model also has to choose among 166 perpetrators. This makes the task of the guessers, either the random or the intelligent version, extremely difficult, and they would be expected to get about 1% of the predictions right.

Importance of Factors

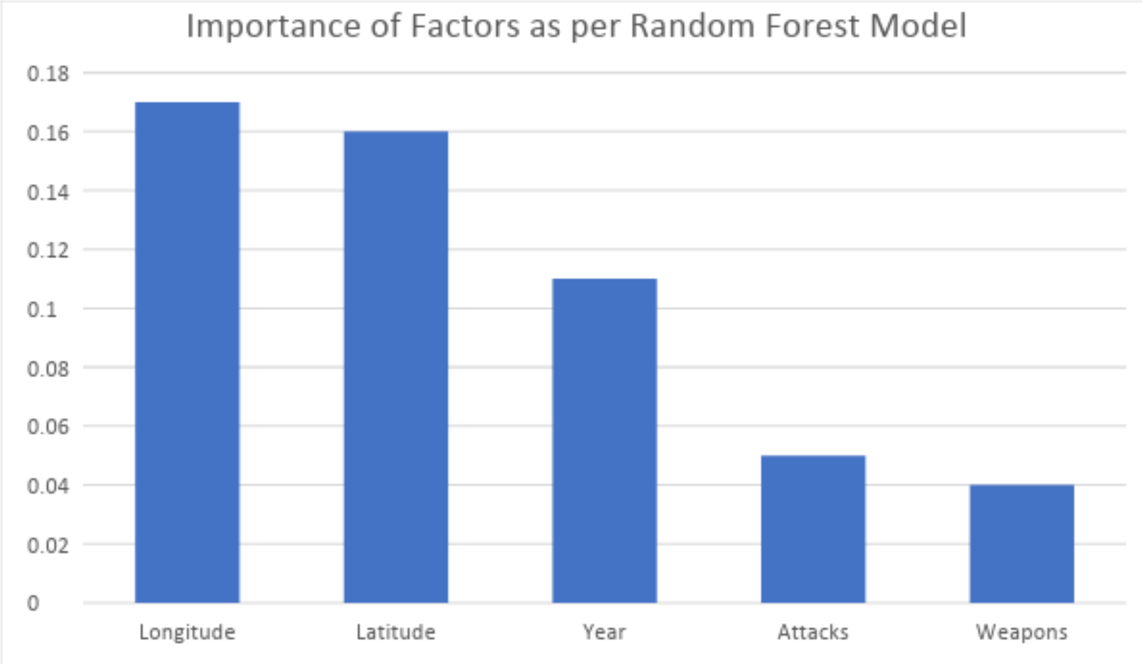
In order to predict the perpetrator, the factors available were the location and year of the incident, the attack, target, and weapons used. The demographic data from the district where the incident happened was also added to the set for analysis to see if that gave any indication into the nature of the perpetrator.



\

We see that the location (latitude, longitude) forms the most important factors with predictive power in identifying the perpetrator followed by the date of the attack. Attacks and Weapons form the next strongest set of predictive power. Demographics factors play a minor role in identifying the perpetrator.

The corresponding importance for random forest model is:



Here also, the location (latitude, longitude, city) is a strong predictor of the identity of the perpetrator group, with attacks and weapons playing a smaller but visible role. The demography and capital had very little predictive value in this model. The year in which the attack happens also had a key role in identifying the perpetrator organization, showcasing the fact that different terrorist organizations are active during different time-periods in the history of India.

Accuracy of Perpetrator Models

When we assess the prediction accuracy of the perpetrator model, the accuracy results are not very good, even when tested on the same training data. This indicates that the data set available may be missing some of the factors that can lead to the identification of the perpetrator group.

When testing on the entire events data after training it on the same, we get the following results over 100 iterations:

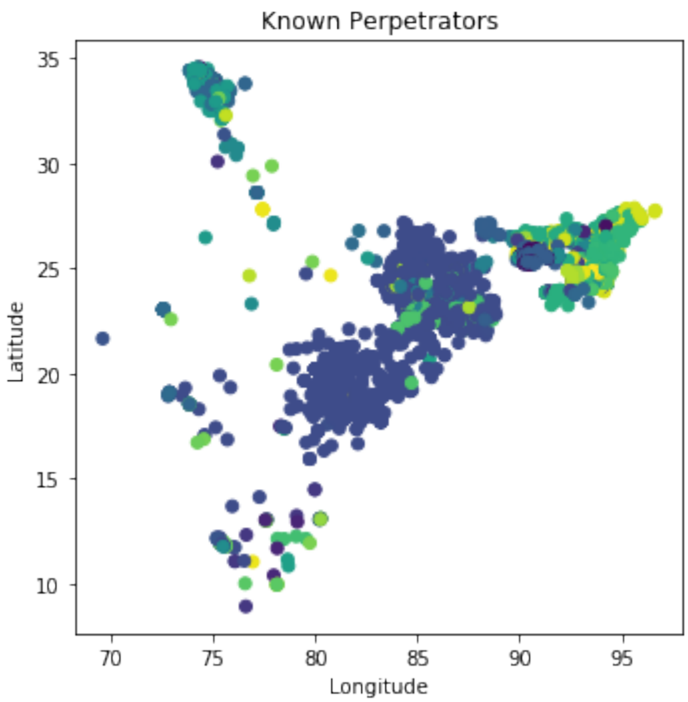
Model	Minimum Accuracy	Maximum Accuracy	Mean Accuracy	Standard Deviation
Decision Tree	97%	97%	97%	0.00
Random Forest	95%	96%	95%	0.2%

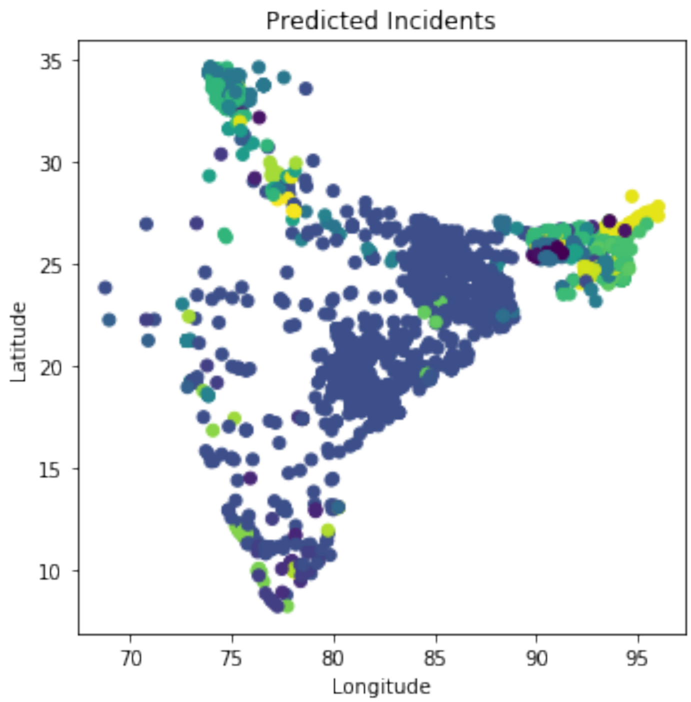
After splitting the event data with 90% for training and 10% for testing, we get the following over 100 iterations:

Model	Minimum Accuracy	Maximum Accuracy	Mean Accuracy	Standard Deviation
Decision Tree	71%	74%	72%	0.7%
Random Forest	68%	75%	71%	1.6%

The 72% accuracy is significantly better compared to random guessing, or even an intelligent guesser of about 1%. Thus, this model can be used to identify the perpetrator for the events where it was unknown in the GTD database. As the reader may recall, the events data consisted of 7697 events of which the identity of the perpetrator organization for only 3744 events was known, and perpetrators for 3953 events were not known.

We used the AI models to predict the identity of the potential perpetrator of the 3953 events. The events with known perpetrators are plotted in the first figure below. The predicted identity of the perpetrators is shown in the second figure below.. The color code used in both figures is the same with each color indicating the same perpetrator. Because of the large number of groups (100+), we are not marking them explicitly with names.





Based on the results of the split training and testing data set, the predictions should be accurate with about 72% probability.

Conclusions:

The following are the answers to the questions posed at the beginning of this document.

1. *Can we predict the important factors that contribute to the occurrence of a terrorist incident in a district?*

The most important factor that can predict this occurrence is its geographic location (latitude and longitude). The other factors that act as markers of the activity is the total population size, population density and the size of the Christian Population, or other non-mainstream religions (e.g. tribal religions). The presence of the capital in the district plays no role in the prediction of whether a terrorist incident will occur.

2. *Can we create an AI model that can predict whether or not a district will have a terrorist incident?*

Yes, we can create such as AI model, which has an accuracy significantly better than an intelligent guesser.

3. *Can we create an AI model that can predict the intensity of terrorism incidents in a district of India?*

Yes, we can create such as AI model, which has an accuracy significantly better than an intelligent guesser.

4. *Can we create an AI model to predict the perpetrator of a terrorism incident?*

Yes, we can create such as AI model, which has an accuracy significantly better than an intelligent guesser.

5. *Can we predict the important factors that can determine the identity of an organization that perpetrated a terrorist incident?*

> The location of the incident, the date of the attack, the type of attack and the weapons used are the most important factors that help identify the perpetrating organization.

Note that the importance of the factors as identified above is based purely on the patterns seen in the available data and should not be considered as causes leading to a terrorist incident. The location (latitude and longitude) may be proxies for other factors that may be causing terrorism incidents, e,g. closeness to the International Border, easy access to sea etc. The correlation of some demographic factors such as Christian Population with occurrences of incidence may be a side-effect of the available data that was examined, e.g. increased influence of communist literature printed in English in areas with Christian Population, where English literacy is higher among the rural population than in the non-Christian communities. Models that we are using can identify correlations, but they do not provide causative explanations (correlation does not imply causation).

POSSIBLE SOURCES

Asal, Victor H. and R. Karl Rethemeyer. (2015). Big Allied and Dangerous Dataset Version 2.0 Available for browsing at http://www.start.umd.edu/baad/database

Alexander, Yonah., and John M. Gleason. 1981. "Behavioral and Quantitative Perspectives on Terrorism." *Pergamon Press*.

Basu, A. 2005. "Social network analysis of terrorist organizations in India." In North *American Association for Computational Social and Organizational Science (NAACSOS) Conference* (pp. 26-28). NAACSOS.

Breiger, Ronald L. 1974. "The Duality of Persons and Groups." *Social Forces* 53(2): 181-190.

Berrebi, Claude., and Darius Lakdawalla. 2007. "How Does Terrorism Risk Vary Across Space and Time? An Analysis Bases on The Israeli Experience." *Defence and Peace Economics* 18(2): 113-131.

Clauset, Aaron., Newman, M. E. J, and Cristopher Moore. 2004. "Finding Community Structure in Very Large Networks" *Physical Review E* 70(6): 1-6.

Crenshaw, M. 2007. "The logic of terrorism." *Terrorism in perspective*, 24.

Everton, Sean. F. 2012. "Disrupting Dark Networks" *Cambridge University Press*

Felmlee, Diane, Lungeanu, Alina, and Derek Kreager. 2017. "Online Dating Preferences: Two-Mode versus One-Mode ERGM Network Analysis." Presented at the meetings of the Population Association of America, Chicago, IL.

Freeman, Linton C. 1979. "Centrality in Social Networks: Conceptual Clarification" *Social Networks* 1: 215-239.

Gartner, Scott Sigmund. 2015. "An Introduction to Net Assessment 2.0: Special Issue on the Net Assessment of Violent Non-State Actors." *CTX: Conflict Terrorism Exchange.*

McCauley, Clark. 2006. "Jujitsu Politics: Terrorism and Responses to Terrorism." In Collateral Damage: The Psychological Consequences of America's War on Terrorism. Westport, CT: Praeger.

National Consortium for the Study of Terrorism and Responses to Terrorism (START). (2016). Global Terrorism Database [Data file]. Retrieved from https://www.start.umd.edu/gtd.

Krebs, Valdis E. 2002. "Mapping Networks of Terrorist Cells" *Connections* 24(3): 43-52.

Newman, M. E. 2008. The mathematics of networks. The new Palgrave encyclopedia of economics, 2: 1-12.

Piazza, James A. 2009. "Is Islamist Terrorism More Dangerous?: An Empirical Study of Group Ideology, Organization, and Goal Structure" Terrorism and Political Violence 21: 62-88.

Piazza, James A. 2009. "Economic Development, Poorly Managed Political Conflict and Terrorism in India." *Studies in Conflict and Terrorism* 32(5): 406-419.

Piazza, James A. 2010. "Terrorism and Party Systems in the States of India." *Security Studies* 19: 99-123.

Reed, Brian. 2007. "A Social Network Approach to Understanding Insurgency" *Parameters* 38: 19-30.

South Asian Terrorism Portal. "India: Terrorist, Insurgent, and Extremist Groups." Online: http://www.satp.org/satporgtp/countries/india/terroristoutfits/index.html

Verma, Dinesh C., Felmlee, Diane, Pearson, Gavin, and Roger Whitaker. 2017. "A Generative Model for Predicting Terrorist Incidents." SPIE.

Wasserman, Stanley., and Katherine Faust 1994. "Social Network Analysis: Methods and Applications*.*" *Cambridge University* Press.

Updated Bibliography

Asal, V., & Rethemeyer, R. K. (2008). The nature of the beast: Organizational structures and the lethality of terrorist attacks. *The Journal of Politics*, *70*(2), 437-449.

Asal, V. H., Rethemeyer, R. K., Anderson, I., Stein, A., Rizzo, J., & Rozea, M. (2009). The softest of targets: A study on terrorist target selection. *Journal of Applied Security Research*, *4*(3), 258-278.

Basu, A. 2005. "Social network analysis of terrorist organizations in India." In North *American Association for Computational Social and Organizational Science (NAACSOS) Conference* (pp. 26-28). NAACSOS.

Berrebi, C., & Lakdawalla, D. (2007). How does terrorism risk vary across space and time? An analysis based on the Israeli experience. *Defence and Peace Economics*, *18*(2), 113-131.

Braithwaite, A., & Li, Q. (2007). Transnational terrorism hot spots: Identification and impact evaluation. *Conflict Management and Peace Science*, *24*(4), 281-296.

Brockhoff, S., Krieger, T., & Meierrieks, D. (2015). Great expectations and hard times: The (nontrivial) impact of education on domestic terrorism. *Journal of Conflict Resolution*, *59*(7), 1186-1215.

Cliff, C., & First, A. (2013). Testing for contagion/diffusion of terrorism in state dyads. *Studies in Conflict & Terrorism*, *36*(4), 292-31

Desilver, Drew., and David Masci. 2017. "World's Muslim population more widespread than you might think." Pew Research Center. Accessed from http://www.pewresearch.org /fact-tank/2017/01/31/worlds-muslim-population-more-widespread-than-you-might-think/.

Diane H. Felmlee and Robert Faris. August 20, 2016. "Toxic Ties: Networks of Friendship, Dating, and Cyber Victimization." *Social Psychology Quarterly*. Vol 79, Issue 3, pp. 243 – 262.

Dinesh Verma, Greg Cirincione, Tien Pham, Bong Jun Ko, "Generation and management of training data for AI-based algorithms targeted at coalition operations," Proc. SPIE 10635, Ground/Air Multisensor Interoperability, Integration, and Networking for Persistent ISR IX, 106350U (4 May 2018).

Enders, W., Sandler, T., & Gaibulloev, K. (2011). Domestic versus transnational terrorism: Data, decomposition, and dynamics. Journal of Peace Research, 48(3), 319-337.

Gartner, Scott Sigmund. 1999. Strategic Assessment in War. New Haven: Yale University Press

Gartner, Scott Sigmund. August 2015. "Net Assessment 2.0: The Net Assessment of Violent Non-State Actors." CTX: Conflict Terrorism Exchange.

Gartner, Scott Sigmund. "Big Data Could Uncover Clue on Marathon." USA Today. Print: 8:A April 17, 2013 (Online 4/16/13).

Gartner, Scott Sigmund and Catherine Langlois. "Unbalanced Policy Priorities and the Interrogation of Terror Suspects." Foreign Policy Analysis, Volume 14, Issue 1, 1 January 2018, Pages 107–126.

Global Terrorism Database. University of Maryland, July 2018, www.start.umd.edu/gtd/do wnloads/Codebook.pdf.

"Kashmir: Why India and Pakistan fight over it." BBC News. Accessed from http://www.bbc.com/news/10537286.

Li, Q., & Schaub, D. (2004). Economic globalization and transnational terrorism: A pooled time-series analysis. *Journal of Conflict Resolution*, *48*(2), 230-258.

Li, Q. (2005). Does democracy promote or reduce transnational terrorist incidents?. *Journal of Conflict resolution*, *49*(2), 278-297.

Marineau, J., Pascoe, H., Braithwaite, A., Findley, M., & Young, J. (2018). The local geography of transnational terrorism. *Conflict Management and Peace Science*, 0738894218789356.

National Consortium for the Study of Terrorism and Responses to Terrorism (START). (2016). Global Terrorism Database [Data file]. Retrieved from https://www.start.umd.edu/gtd.

Piazza, J. A. (2006). Rooted in poverty?: Terrorism, poor economic development, and social cleavages. *Terrorism and political Violence*, *18*(1), 159-177.

Piazza, J. A. (2008). Incubators of terror: Do failed and failing states promote transnational terrorism?. *International Studies Quarterly*, *52*(3), 469-488.

Piazza, James A. 2009. "Is Islamist Terrorism More Dangerous?: An Empirical Study of Group

Ideology, Organization, and Goal Structure" Terrorism and Political Violence 21: 62-88.

Piazza, James A. 2009. "Economic Development, Poorly Managed Political Conflict and Terrorism in India." *Studies in Conflict and Terrorism* 32(5): 406-419.

Piazza, James A. 2010. "Terrorism and Party Systems in the States of India." *Security Studies* 19: 99-123.

Piazza, J. A. (2011). Poverty, minority economic discrimination, and domestic terrorism. *Journal of Peace Research*, *48*(3), 339-353.

Piazza, J. A. (2013). Regime age and terrorism: Are new democracies prone to terrorism?. *International Interactions*, *39*(2), 246-263.

"Population Census 2011." Religion Data - Population of Hindu / Muslim / Sikh / Christian -Census 2011 India, Census Organization of India, 2011, www.census2011.co.in/.

"Religion Census 2011." Religion Data - Population of Hindu / Muslim / Sikh / Christian -Census 2011 India, Census Organization of India, 2011, www.census2011.co.in/religion.php

Rithvik Yarlagadda, Diane H. Felmlee, Dinesh Verma, Scott Sigmund Gartner. (2018) Implicit Terrorist Networks: A Two-Mode Social Network Analysis of Terrorism in India. In: Thomson R., Dancy C., Hyder A., Bisgin H. (eds) Social, Cultural, and Behavioral Modeling. SBP-BRiMS 2018. Lecture Notes in Computer Science, vol 10899. Springer.

Saxena, S., Santhanam, K., & Basu, A. (2004). Application of social network analysis (SNA) to terrorist networks in Jammu & Kashmir.*Strategic Analysis*, *28*(1), 84-101.

South Asian Terrorism Portal. "India: Terrorist, Insurgent, and Extremist Groups." Online: http://www.satp.org/satporgtp/countries/india/terroristoutfits/index.html.

Wilson, M. C., & Piazza, J. A. (2013). Autocracies and terrorism: Conditioning effects of authoritarian regime type on terrorist attacks. *American Journal of Political Science*, *57*(4), 941-955.