

CAPSTONE PROJECT PROPOSAL 1

What is the problem you want to solve?

Classifying and predicting if terrorist attacks are perpetrated by domestic or foreign nationals in the US

Who is your client and why do they care about this problem? In other words, what will your client DO or DECIDE based on your analysis that they wouldn't have otherwise?

Terrorism is open to different interpretation. Depending on who you ask, the responses you get would be different. Most will probably say it is a negative foreign impact on lives and properties. However, recent incidents like in the case of Las Vegas shootings on October 1, 2017, and that of Texas church shootings a month after tells a different story. Could such be classified as terrorist attacks? Are such perpetrated more by domestic or foreign nationals based on the history of similar occurrences? Can we sort and predict present or future similar incidents? These questions form the core of my analysis? It is essential to anyone that cares about safety to lives and properties, in that it helps:

- the government, know where to concentrate homeland security efforts.
- businesses - large and small, to estimate security risk and plan mitigation measures
- residents in the US – citizens and non-citizens, to change the perception of causes of such occurrences, and to keep guard in their daily endeavors.

What data are you going to use for this? How will you acquire this data?

I would be using the imbalanced dataset from Global Terrorism Database ([GTD](#)) maintained by the University of Maryland.

In brief, outline your approach to solving this problem (knowing that this might change later)?

- Run a hypothesis test on the dataset to check if the difference between the attack proportions of citizens and foreigners is barely by chance or highly significant to be considered such at a significant level of 0.05.
- Using the Random Forest algorithm in Python on the already preprocessed data to classify and predict if domestic or foreign nationals perpetrate such, and also checking for performance indicators of the model.

What are your deliverables? Typically, this would include code, along with a paper and/or a slide deck.

Deliverables will be code and paper explaining my analysis in details.