Capstone Project Proposal

Josh You

For my first capstone project, I will analyze how the characteristics of an attempted terrorist attack affect the likelihood that the attack will succeed.

My client for this project would be government agencies that seek to promote security and prevent successful acts of terrorism. The outcome of this project may determine what types of individuals and organizations pose the most concerning security threats, which would affect how agencies allocate attention and resources in investigating terrorism threats, in addition what types of weapons and techniques affect the likelihood of an attack succeeding, which may inform regulatory approaches to preventing violence.

The data for this project will come from Global Terrorism Dataset compiled by the National Consortium for the Study of Terrorism and Responses to Terrorism (START), which is a research center based in the University of Maryland. The dataset contains information on 170,000 attacks between 1970 and 2016. The dataset is publicly available at https://www.kaggle.com/START-UMD/gtd.

The dataset contains over 100 variables describing various characteristics of each attack. The main outcome of interest is success, which is a binary categorical variable. The dataset also contains information on the number of people killed and injured in the attack, and the extent of property damage. Other variables of interest include the location, attack type (including assassination, armed assault, bombing, hijacking), weapon type, characteristics of the target, and many more.

My approach for this project will be to use the data set to train a classifier to predict whether an attack will succeed or not based on its characteristics. I will likely experiment with several different machine learning techniques. I may also use regression to explore the relationship between an attack's characteristics and the numeric results (casualties, property damage).

My deliverables for this project will be a paper summarizing and discussing my findings, as well as any Python code used to analyze the data and produce the results.