Proposal for Capstone Project

Title:

Predicting the next Newtown or Parkland shooting

Problem:

A leader of the free-world, United States of America has more people dying of gun violence than any other developed country, 29.7 per 1 million people. Since Sandy hook, Newtown incident in 2012, there have been over 1600 mass shootings in America. A mass shooting is defined as over 3 deaths in a shooting, excluding the gunman. Where there is no doubt that more guns equal more deaths due to guns, the debate on gun violence intensifies with an Parkland shooting like incident, but lasts for just a few weeks resulting in no significant changes in the gun laws. As a matter of fact the rate of mass shooting has increased in the last 2 years as compared to the last 6 years combined. And with another political standstill, there will be more victims to the mass shootings in America.

Who might care?

Anybody with a right mind in America should care about and value human life. To minimize the problem, changing laws at a federal level alone will not solve the problem. This study can enable the local law enforcement, medical, and government agencies to reduce the possibility of a mass shooting in their community. This project can bring significant changes to today's approach to gun violence, from reactionary to precautionary.

Data:

The data will be acquired from the following

1. https://cis.upenn.edu/~ccb/publications/gvdb-d4gx.pdf
2. https://www.statista.com/topics/1287/firearms-in-the-us/
3. http://www.gunviolencearchive.org/reports
4. https://crime-data-explorer.fr.cloud.gov/
5. https://www.nami.org/learn-more/mental-health-by-the-numbers
6. https://www.motherjones.com/politics/2012/12/mass-shootings-mother-jones-full-data/

The data contains information of gun related incidents from early 90's to present by year, state, city. The data set includes, data on crime, mental health and gun ownership by state, city, year, gender and age. For this project, I will focus on 15 years of data, starting year 2002, with an underlying assumption that the data prior to 2002 is not going to significantly impact this analysis, due to changes in geo-political, cultural and technological trend and outlooks.

Modeling approach:

I will explore the correlation between gun-sales, mental-health, general crime, and mass shootings. TBD…. Need to know more about the modeling techniques, algorithms etc….

Read:

1. https://www.wired.com/story/crime-predicting-algorithms-may-not-outperform-untrained-humans/
2. https://science360.gov/obj/video/bd241f71-ccc8-4aee-a819-e249345bbbe4/predpol-algorithms-predicting-crime

Possible limitations:

Unfortunately, less than 5% of the mass shootings ever get any national media attention. In the other 95%, the factors leading to "mass shootings" is family or personal disputes. Which are very challenging to ascertain and predict. Given my current skillset and other obvious challenges, I have currently, also left-out the social media analysis for this project. I might reconsider this at a later stage in the project.

Deliverables:

1. Codes (notebooks) for:
   a. data acquisition
   b. data cleaning
   c. data exploration analysis
   d. machine learning model development

2. Report on the capstone project
3. Presentation on the capstone project