DATA 698

Capstone Project Proposal

September 16, 2022

Prepared for:

Prof. Dr. Nasrin Khansari
City University of New York, School of Professional Studies

Prepared by:

Sie Siong Wong
Mario Pena
Joseph Shi

Many of us spend several hours per week in social media platforms such as Facebook, Twitter, and Instagram to name a few. Although there are many advantages that social media provides e.g. serving as the fifth estate of power, a place to express opinions and such, it has also become a platform for someone to spread hate, cyberbullying, harassments and so on. Messages of hatred and spread of violence could have a high potential for influencing and motivating someone to commit a crime. According to CNN, hate crimes in New York City have increased 76% compared to last year. Many social media companies such as Facebook, have started to control the contents being posted by their users. Because there is a big concern in our society that offensive speech can result in more crimes and eventually become a serious threat to social, political, and cultural stability, we see the need to find a solution.

Some may argue that the First Amendment allows individuals to express anything they want. Instead of limiting everyone's freedom of speech openly in social media platforms, there are possible alternatives to deal with this issue. For example, a platform can allow users to turn on or off a switch to hide possible sensitive contents. In order to do this, we need some kind of solution that is able to identify inappropriate speech, tag them, and warn users that their post may not be visible to everyone due to containing sensitive language.

We believe that offensive speech has a high potential for causing more violent crimes to happen in the city. To make our city a better place to live, we need to find a way not only to identify inappropriate speech in social media platforms and reduce the spread of such speech, but also when to deploy more law enforcement to patrol areas crowded with high numbers of offensive speech in the hope we interrupt crime prematurely. Our research hypothesis is that higher offensive tweets from an area have strong correlation to the number of crimes committed in that area and minority groups are mostly the target.

The challenging part for our research is to classify the tweets metadata into offensive speech. Defining what is offensive speech such as hate speech can be hard because it's constantly evolving and often dependent on context. Until today, artificial intelligence still struggles when it comes to identifying hate speech. But we've found that quite a few research scientists have tried different approaches to identifying offensive speech in social media text data. The more advanced techniques, deep learning methods such as gated recurrent unit (GRU), convolutional neural networks (CNN), gated convolutional recurrent - neural networks (GCR-NN), and combination of transfer learning and week supervision, are so far giving the best results compared to conventional classification approaches such as logistic regression, naive Bayes, decision tree, random forest, and gradient boosting.

There are many different types of hate speech that target different groups and individuals, and research has been done to differentiate and define them accordingly. Some approaches to the classification of hate speech from Tweeter use the hashtag as a variable for predicting violence and hate message from the tweet. While other approaches for classification of hate speech on social media may also include the development of neural language models. Additionally, there are algorithms that have been created to detect hate speech in digital microenvironments, which facilitate and reduce analysis tasks undergone by law enforcement agencies and service providers. We have also come across research that found there may not be a correlation between social media messages and crime, and others that consider "weather" to be an important environmental factor that affects the occurrence of criminal incidents. There have been many different approaches tried to find a link between social media posts and crime, all with varying results, and what makes this research so interesting is that we may yet find different results that may contribute to this solution.

There are two main parts in our study. First part is to identify offensive speech from tweets metadata. Then, we will use the finalized offensive tweets data, which have the date and coordinates link to the crime data reported in NYC. We have chosen to use crime data from year 2020 and on because since the pandemic outbreak until today, there was a sharp rise in crime compared to prior years. Likewise, we will collect NYC tweets data from year 2020 to date by using Twitter API. We have done the data collection through Twitter API and NYC Open Data API. There was no limitation to read crime data from the NYC Open Data source (NYPD Complaint Data Historic dataset) but for Twitter API there was. Therefore, we applied for an academic research access account, which has a higher tweets cap and we were approved. We'll use a sentiment analysis approach to annotate tweets into positive, negative, and neutral sentiments. Then, we further classify all negative sentiment tweets into hate or offensive speech by using either topic modeling or refer to a collection of defined offensive speech. All identified offensive tweets will then be mapped to the crimes reported by using date and coordinates for correlation and time series analysis. We'll focus our

study in the five boroughs of New York City: Manhattan, Brooklyn, Queens, Bronx, and Staten Island. For performance comparison, several well-known classification machine learning models such as random forest, KNN, SVM, will be implemented using the optimized parameters and a deep learning model. Recurrent neural networks (RNN) will also be considered. Below is the architecture of the proposed methodology.

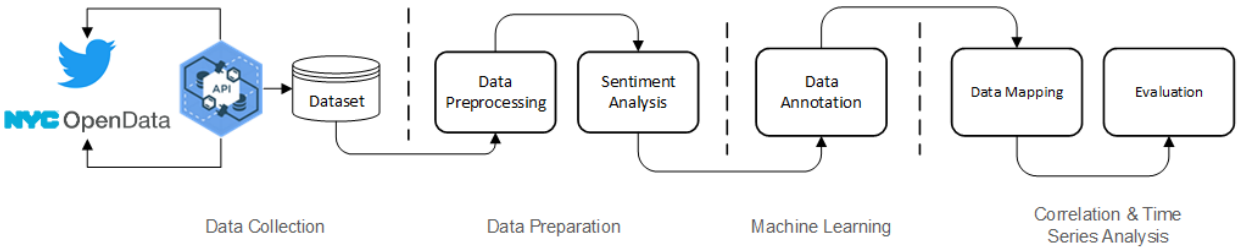

Figure 1: Architecture of the Proposed Methodology