### **CMPT 732: Big Data Project Proposal**

### **Analysis of Hate Crime Data in United States of America**

Arin Ghosh

Muhammad Muhaimin

Karanjit Singh Tiwana

Hatred can be spotted throughout the history of civilization. People have become victims of hate crime based on their race, gender, sexual orientation or belief. In the internet era, hatred is merely binary numbers that float across cyberspace but indeed it can be overtly devastating. The current outbreaks that occupy social media and the internet as a whole have signified to us that we are living in a society oiled by hatred.

This project aims to work on an exploratory analytics tool that might help us to dig the facts behind the deeply spreaded roots of hate crime and give us an understanding of the sources behind it and the trend it follows. We are planning to use the open source data provided by the platforms like National Archive Of Criminal Justice Data, Kaggle and FBI for our project.

Plan:

The project is divided into two parts. In first part, our plan is to start exploring the hate crime data given by the FBI: <https://ucr.fbi.gov/hate-crime/> and [National Archive Of Criminal Justice Data.](https://www.icpsr.umich.edu/icpsrweb/NACJD/series/57/studies?searchIn=TITLE&archive=NACJD&q=%22Hate+Crime+Data%22&sortBy=7)

Both of the sites have archived hate-crime related data on a year to year basis. Each of the year like this one https://ucr.fbi.gov/hate-crime/2015. It has a ton of data scattered across different tables like

* Incidents and Offenses
* Victim
* Offenders
* Location Type
* Jurisdictions etc.

In the later part of the project, we are planning to tie the FBI data set with the political affiliation for a particular state. Once we incorporate the political data into our existing analysis we may bring forth new information with something like the hate crime in any state has increased over time due to political influence or not. Especially it will be interesting to find out these answers for *‘swing states*’.

The core of our project is to draw various graphs and use data science methods to display the story or the trend behind this seemingly uncorrelated data. We will be doing most of our parsing and cleaning of data using Spark and for presentation/UI we are planning to use R’s Shiny server or Tableau public. The final polished UI will serve as a tool to find out relationships between different elements of the FBI hate crime data. The graphs & the tables will be interactive and the filtered result from the tool can be downloaded as well for further analysis.

Future plans:

If time permits we also plan to extend the project by analyzing how fake news affected hate crimes. But we are yet to figure out the timeline for our initial project plan. Therefore depending on the progress and the time left in our hand before the demo, we may give this a try. But as of now, we are not 100% sure about if this is doable within the project deadline day.