Minneapolis Police Data Analysis

Brennan Rhoadarmer Nithish Reddy Kolby Johnson Fateh Sandhu

University of Nebraska - Lincoln, Computer Science Department

—————————— ◆ ——————————

# INTRODUCTION

For this project we wanted to do some analysis over police use of force data and police stop data in a major metropolitan area. To augment the police data, we will be examining and analyzing alongside precipitation data from the same areas. This analysis can be used to determine if there is bias in police operations or if there are times, places, or weather conditions that cause more crime or police action to occur.

# OBJECTIVES

Our objectives for this assignment are as follows:

* Preprocess the minneapolis police data and the weather data, combining the two into a single data set and dealing with incomplete or missing data.
* Use nominal data to do association analysis in police stop/crime data.
* Do spatial clustering analysis with visualization
* Conduct temporal-spatial cluster analysis with visualization.

# RELATED WORK

Most of the work is to analyze the patterns and trends in the crime. By using data mining techniques to make previously unknown crime data to useful data to observe the technologies used by criminals and help police to use advanced technologies, to put police ahead of criminals[4].

# PROBLEM DEFINITION

For this project we are analysing police action data and working to derive a pattern in either the suspected criminals that they are interacting with, or within the actions of the police. Therefore, our input will be those police actions while our outputs will be different cluster, frequency and association analysis that will allow us to determine if such patterns exist, leading us to find why those patterns exist. We may eventually focus on different key events that may change patterns that we normally see such as protests, riots, or other significant incidents.

# DATASET

The main datasets that we are using are police stop and police use of force data provided by the Minneapolis government website. The police stop data has incidents from October 31, 2016 until present day and is updated daily and contains 173.6k reports while the police use of force data has reports from January 2, 2008 until present day with 32k reports and is also updated daily. To pair with this, we are also utilizing weather data from CoCoRaHS which has weather data from 2016 until present, with 500,000 entries for all across Minnesota. All of the datasets have date and time, along with latitude and longitude.

# APPROACH

For our analyses, we will start with simple methodology that we already have implementations for, such as k-mean or apriori algorithms. We will then transition to using more complex tools for analysis and cross check each result we get with previous results to see if they line up, or give contradictions.

For software specifically, we plan on using Python as our main programming language, and utilizing NumPy, MatPlotLib and Pandas as third party libraries to assist us in our analysis. We are going to use Jupyter notebooks as staging grounds for the actual analyses, and using git for version control.

We will also be utilizing some abilities of Weka in order to check our work, and to ensure that we derive some results.

# EVALUATION

Part of our evaluation will be dining by comparing results that we get from our programming and analyses with what Weka gets from its own algorithms and analyses. Some of our algorithms will also have outputs of an expected format or range, and so if we do not get results that match our expectations we know to look further into our implementation. Finally, our visualization of our results will help us see directly, if any of our analyses do not make sense, or have returned grossly wrong results.

# IMPLEMENTATION PLAN AND TIMELINE

Most of the software that we are going to be using was listed in the approach section.

For the timeline for the project: In week one, we will work on the preprocessing of data, completing objective one. This will be complete by the end of day, March 28th so that we can move onto the objectives that will allow for analysis. Objective two will then be started, hopefully finishing both the program, analysis and presentation for it byApril 11th, with Objective three following, competing on April 25. This leaves one and a half weeks for Objective four, which as more of an extension of Objective 3, should be possible if our timing is right and we can push ourselves. Then we should have one or two days to compile our findings and code and finish up the report.

# REFERENCES

1. [Police Stop Data](https://opendata.minneapolismn.gov/datasets/police-stop-data)
2. [Police Use of Force](https://opendata.minneapolismn.gov/datasets/police-use-of-force)
3. [Community Collaborative Rain, Hail & Snow Network](https://www.cocorahs.org/ViewData/StateDailyPrecipReports.aspx?state=MN)
4. C. Chauhan and S. Sehgal, "A review: Crime analysis using data mining techniques and algorithms," 2017 International Conference on Computing, Communication and Automation (ICCCA), Greater Noida, India, 2017, pp. 21-25, doi: 10.1109/CCAA.2017.8229823.