**Project Proposal**

CSP 554 Big Data

Group Name: Chicago Crime Analysis

Group Members: Radhika Barbole, Shivam Kulkarni, Shubham Jain, Vivekanandan Sakthivel

Project Description: Details of all crimes that happened in the city of Chicago since 2001 are recorded to create the dataset which we would like to work with for this project. The dataset has 6.99 million records, 30 attributes and the file size is 1.7 GB. Our aim is to analyze the data and derive some useful insights from it. To achieve our aim, we are planning to use several Big Data tools like HDFS, Hive, Pig, and Spark. For Machine Learning algorithms we will use Spark Machine Learning Library (MLlib).

Question we are going to address by using Hive, Pig:

1. What are the most occurring crimes in the city?

2. How many crimes are being committed at a specific location? (e.g. Street, residence)

3. Which crimes are being committed at a specific time of the day?

4. At what locations are the crimes being committed at a specific time of the day?

5. Which crimes are being committed at a specific day of the week?

6. At what locations are the crimes being committed on a given day of the week?

7. Analysis of a particular crime type over the years

For visualizations, the results of the questions 3,4,5,6 will be represented by heatmaps.

Question we are going to address by using Spark MLlib:

1. What kind of a crime random person is likely to face at a given location, date and time? (Prediction)

2. What types of crimes are more likely to happen in which part of the city? (Clustering)

References:

Research Papers:

~~[1] Nathan Holt Analyzing Crime in Chicago Through Machine Learning, available at:~~ [~~http://nathanwayneholt.com/mathematicalmodeling/ChicagoCrimesReport.pdf~~](http://nathanwayneholt.com/mathematicalmodeling/ChicagoCrimesReport.pdf)

[1] Mcclendon, Lawrence, and Natarajan Meghanathan. “Using Machine Learning Algorithms to Analyze Crime Data.” *Machine Learning and Applications: An International Journal*, vol. 2, no. 1, 2015, pp. 1–12., doi:10.5121/mlaij.2015.2101.

[2] Shyam Varan Nath. 2006. Crime Pattern Detection Using Data Mining. In Proceedings of the 2006 IEEE/WIC/ACM international conference on Web Intelligence and Intelligent Agent Technology (WI-IATW '06). IEEE Computer Society, Washington, DC, USA, 41-44. DOI: <https://doi.org/10.1109/WI-IATW.2006.55>

[3] Kim, Suhong & Joshi, Param & Kalsi, Parminder & Taheri, Pooya. (2018). Crime Analysis Through Machine Learning. 415-420. 10.1109/IEMCON.2018.8614828. Available at: <http://airccse.org/journal/mlaij/papers/2115mlaij01.pdf>

Dataset:

[4] Crimes - 2001 to present | City of Chicago | Data Portal. url: <https://data.cityofchicago.org/Public-Safety/Crimes-2001-to-present/ijzp-q8t2>