Prakash Wagle Aditya Bhatkar

6190 Course Project Proposal

Urban Crime Analysis

Problem Statement

This project aims at providing urban crime analysis. The dataset provides historic data criminal offenses in the San Francisco city. The project will be implemented in Apache Spark.

Inspiration

With the increasing wealth inequality, limited resources and rising aspirations the crime in cities is also increasing rapidly. This makes citizens feel unsafe and cops busy. But there might be a pattern in the way the crimes take place. This project attempts to find out those patterns in order to tackle the crime efficiently.

Achievable goals

Clustering of dangerous city neighborhoods: K-Means clustering

Identifying frequent types of crime: Frequency Count

Classification of new crime category: K-NN classification

Generate reports on data: Hive Queries

Likely to accomplish

Visualizing neighborhoods and the crime category at the given time and day

Ideal accomplishment

The system will be able to advice user on the risk level in a neighborhood at a given time and day: Logistic regression.

Dataset description

The dataset is taken from one of the problems in kaggle competitions.

https://www.kaggle.com/c/sf-crime/data?train.csv.zip

Data fields

Dates - timestamp of the crime incident

Category - category of the crime incident (only in train.csv).

Descript - detailed description of the crime incident (only in train.csv)

DayOfWeek - the day of the week

PdDistrict - name of the Police Department District

Resolution - how the crime incident was resolved (only in train.csv)

Address - the approximate street address of the crime incident

X - Longitude

Y – Latitude

Project Contributors

1. Prakash Wagle 800889950
   Analyst and Developer
2. Aditya Bhatkar 800887086
   Analyst and Developer