**Project Proposal**

**Description of the Dataset:**

The main idea of the project is to develop a data mining application which is based on citizen complaint analysis. The data used for this application is a citizen complaint data of Detroit for the year 2016. This data is provided by the Detroit Police Department on the open data portal for the city of Detroit.

The results of this application would be helpful in detecting areas with more complaints from the citizen.

The data consists of 2536 records with 13 attributes. The attributes contain the BPC, CCR, Report date, Entry, Age, citizen race, citizen sex, complaint closed date, Unit, allegation, Finding, officer race, officer sex. BPC stands for Board of police commission. Report date is the date when the citizen lodged the complaint. Entry is type way how the citizen lodged the complaint. Unit is the department of police where the complaint was forwarded. Allegation is description made by the citizen and finding is the actual finding of the situation by the police. The officer sex and office race are also provided in the data set.

|  |  |
| --- | --- |
| Name of the Dataset | DPD Citizen Complaints |
| URL | https://data.detroitmi.gov/datasets/dpd-citizen-complaints |

**Problem Statement:**

Governments need to consider citizen’s satisfaction an important factor and provides means for to citizen to report complaints. Data mining of the citizen’s complaints data provide the police department and the government to evaluate the citizen’s needs. This project will analyse the citizen complaints data collected by the Detroit Police Department using clustering and classification techniques it is possible to find primary causes for complaints.

**Data Mining Approach:**

To detect the type of complaints, clustering can be used to cluster different types of complaints provided in the data set. Different data mining parameters like classification, clustering, visualization, prediction can be applied on this data to generate various conclusions.

**Data Transformation:**

Data pre-processing contains of several techniques that are useful to pre-process the data before data mining. Data pre-processing also includes techniques to handle the missing data values from the data set. There are different ways to handle missing data values. Data pre-processing also includes techniques such as sampling the data, dimensionality reduction, feature creation, feature subset selection, Normalization and Standardization. Some techniques need only categorical data and some need only nominal data. Discretization can be used to discretise the data so that the numeric values are converted in to the nominal values. In this project, we will be using discretization to discretise the age data as there are 73 distinct values of age for better classification and clustering.