VIT®

Vellore Institute of Technology

(Deemed to be University under section 3 of UGC Act, 1956)

SCHOOL OF COMPUTER SCIENCE AND ENGINEERING (SCOPE)

November 2019

CSE3020 - DATA VISUALISATION

FINAL PROJECT REPORT

CHICAGO CITY CRIME VISUALISATION

Under the guidance of,

Prof. ANNAPURNA JONNALAGADDA

Submitted By:

Varun Kaushik (17BCE0182)

Pulkit Gupta (17BCE0631)

SLOT: G1

PROBLEM STATEMENT

Why is this dataset more important now? Because the number of Crime in Chicago increase, more and more people care about their safety, and the government leader want to build a good environment for the citizens. Therefore, it is required to predict the occurrence of a crime at a location at a specific time of a day and to anticipate if a particular person in the city, at any given duration of the day will be a crime hotspot or not, with an acceptable rate of accuracy. This dataset reflects reported incidents of crime (with the exception of murders where data exists for each victim) that occurred in the City of Chicago from 2001 to present, minus the most recent seven days. Data is extracted from the Chicago Police Department's CLEAR (Citizen Law Enforcement Analysis and Reporting) system. In order to protect the privacy of crime victims, addresses are shown at the block level only and specific locations are not identified.

INTRODUCTION

According to the Chicago Crime report records, “Crime in Chicago has been tracked by the Chicago Police Department’s Bureau of Records since the beginning of the 20th century. The city’s overall crime rate, especially the violent crime rate, is substantially higher than the US average. Chicago was responsible for nearly half of 2016’s increase in homicides in the US”. Chicago’s homicide rate is higher than the larger American cities of New York and Los Angeles, the reasons for the higher numbers in Chicago remain unclear However, “the Chicago police department tallies data differently than police in other cities, the FBI often does not accept their crime statistic

Keeping these concerns in mind, we have taken the dataset to analyse and perform prediction on crime events that occurred in the City of Chicago from 2001 to 2017. For our project, we have more than 6280882 millions of the data records. The research aim is to delve deeper into this statistic and deal with missing values in our dataset. We use several graphs to analyse the dataset.

SIGNIFICANCE AND APPPLICATION

With the access to large amounts of data and with increasingly smarter statistical analysis, we will be able to use the collected data to foresee and percept various types of criminal acts at a particular time and location. Such insight will enable city authorities to detect areas of increased crime, which will give them the opportunity to act more deliberately and deploy officers more intelligently, e.g. sending them to the areas that are more exposed to crime. With such systems, police officers will not only respond to criminal acts, but will also be able to act proactively and stop them before they occur.

LITERATURE SURVEY

[1] Kassen, M. .One of the most important aspects of spatial crime analysis is the identification of hotspots: areas of the highest crime concentration. This paper advances a methodology for hotspot detection based on a global moving window approach combined with the use of local statistics to define the hotspot limit

[2] Nakaya, T., & Yano, K. .In this paper they propose for an effective interpretation of spatio-temporal patterns of crime clusters/hotspots, they explore the possibility of three-dimensional mapping of crime events in a space-time cube with the aid of space-time variants of kernel density estimation and scan statistics.

[3] Ratcliffe, J. H. . The aim of this paper is to broaden this debate out to consider the effectiveness of a wider range of visualisation techniques in permitting an understanding of spatio-temporal trends. The merits of three visualisation techniques, (map animation, the comap and the isosurface) are evaluated on their ability to assist in the exploration of space–time patterns of crime disturbance data. They conclude that each technique has some merit for crime analysts charged with studying

[4] Shiode, S., Shiode, N., Block, R., & Block, C. R. , Crime analysis and prevention is a systematic approach for identifying and analyzing patterns and trends in crime. In this paper they have an approach between computer science and criminal justice to develop a data visualisation procedure that can help solve crimes faster.

[5] Aurisano, J., Kumar, A., Gonzales, A., Reda, K., Leigh, J., Di Eugenio, B., & Johnson, A. ,This paper seeks to make a contribution. One of the newest tools in this area is crime mapping using Geographical Information Systems (GIS The aim of this paper is to compare a couple of different ways of visualising high volume crime data and then to discuss some, possibly unforeseen, implications of these mapping techniques

[6]] Goldschneider, M. , This investigates patterns of micro-scale concentrations of different types of crime using the network distance in the spatial, temporal and spatial-temporal dimensions. The result of this paper demonstrates the significance of the street-level analysis from the microscopic perspective, which can help form a more focused policing tactic.

[7] Rossy, Q., & Ribaux, O. ,This paper mainly focuses on visualising the data so that is understable by the people. Visual data exploration poses challenges for 'InfoVis Novices'. A 'conversational interface' which would enable users to generate and interact with visualizations through natural language and gestures, while maintaining a history of the data exploration, has the potential to ameliorate many of these challenges

[8] Brunsdon, C., Corcoran, J., & Higgs, G. This article presents a case study of the open data project in the Chicago area. Open data is a concept that governmental data should be available to anyone with a possibility of redistribution in any form without any copyright restrictions.

[9] Gottschalk, P., & Tolloczko, P. C. The article discusses the issue of digital technologies use for practical applications of the principles of modern ecologic currents in criminology. So they proposed more advanced studies on spatial distribution of crime attempting at explaining this phenomenon.

[10] Ratcliffe, J. H., & McCullagh, M. J. .This paper talks about the visualisation of the crime data, Indeed in a collaborative framework, different kinds of visualisations forensic case data can play a central role for supporting decisions

[11] Rauscher, J., Swiezinski, L., Riedl, M., & Biemann, C. This paper develops a maturity model for geographic information systems in crime mapping. Their purpose of such a model is to help practitioners and researchers study organisational evolution and determine future direction in police organisations' use of electronic systems when mapping crime.

[12] Bonatsos, A., Middleton, L., Melas, P., & Sabeur, Z.They present CoocViewer, a graphical analysis tool for the purpose of quantitative literary analysis, and demonstrate its use on a corpus of crime novels. The tool displays words, their significant co-occurrences, and contains a new visualization for significant concordances.

[13] Vasiliauskas, D., & Beconytė, G. This paper describes the major research and development activities which have been achieved so far since the launch of the DESURBS project. It allows users to explore historical crime trends for a region over time, where crime statistics are contrasted.

[14] Brunsdon, C., & Corcoran, J. This paper mainly deals with the Cartography of a crime which is mainly used in Europe, especially in Eastern Europe, geographic research in criminology deals. Paper proposed that the set of maps, which were designed through cartography, represent specific characteristics, density and temporal distribution of crimes in the city.

[15] Oatley, G. C., & Ewart, B. W. They begin the paper by describing circular statistics. A set of techniques referred to as circular statistics has been developed for the analysis of directional and orientational data. They then discuss how these may be modified, and demonstrate the approach with some examples for reported incidents in the Cardiff area of Wales.

[16] Sathyadevan, S., & Gangadharan, S. The Project was developed primarily to assist the Police with the high volume crime, burglary from dwelling houses. A developed software system enables the trending of historical data, the testing of 'short term' hunches. The software utilises mapping and visualisation tools and is capable of a range of sophisticated predictions.

[17] Ratcliffe, J. H. This paper demonstrates a technique that uses police START and END crime times to generate a crime occurrence probability at any given time that can be mapped or visualized graphically. Police crime data tends, more often than not, to reflect the routine activities of the victims rather than the offense patterns of the offenders.

[18] Corcoran, J., Wilson, I. D., Lewis, O. M., & Ware, J. A. .Crime rates differ between types of urban district, and these disparities are best explained by the variation in use of urban sites by differing populations.This paper outlines the first stage in the development of a system, designed to facilitate the prediction of crime hot spots. For this stage, a series of Kohonen Self-Organising Maps (KSOM) will be used to cluster the data in a way that should allow common features to be extracted.

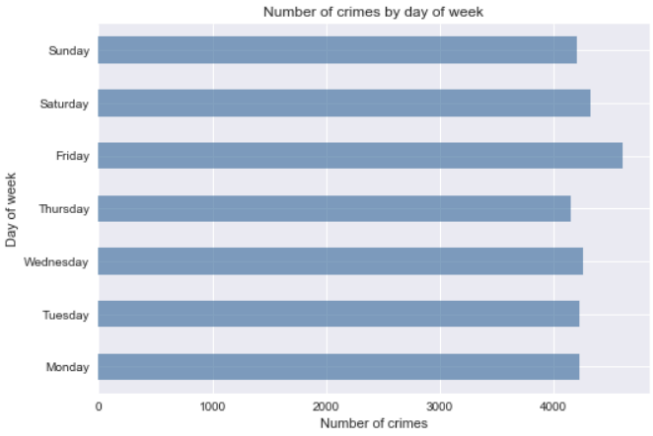
[19] Adderley, R. W., & Musgrove, P. This paper provides an overview of the role computer software plays within police forces with particular attention paid to crime analysis and investigation computer systems. Illustrative systems that are in practical use for tackling both major and volume crime are described. Particular attention is paid to the attempts that have been made to apply artificial intelligence techniques to tackling the volume crime of burglary.

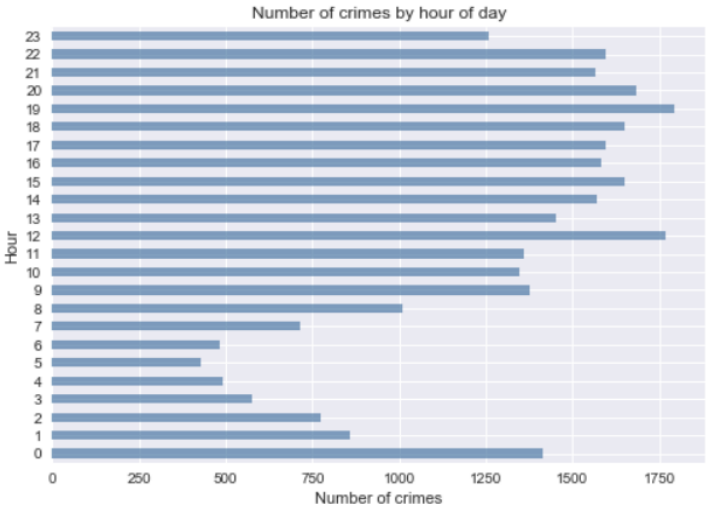
[20] Kang, H. W., & Kang, H. B. This paper talks about the visualisation of the crime data, Indeed in a collaborative framework, different kinds of visualisations integrating forensic case data can play a central role for supporting decisions. Among them, link-charts are scrutinised for their abilities to structure and ease the analysis of a case by describing how relevant entities are connected.

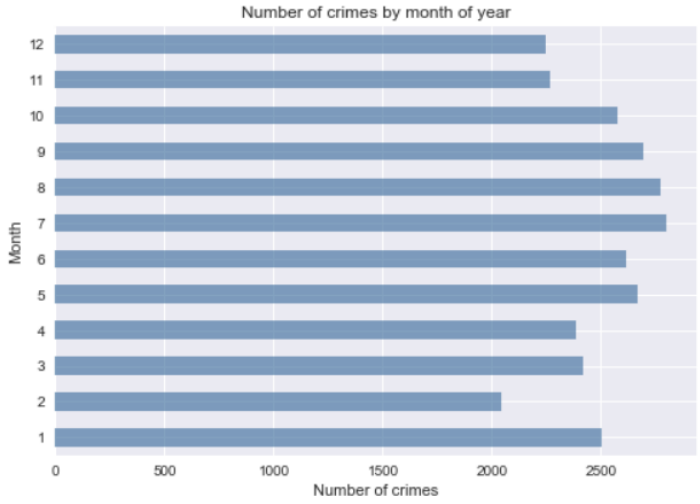
Proposed Work

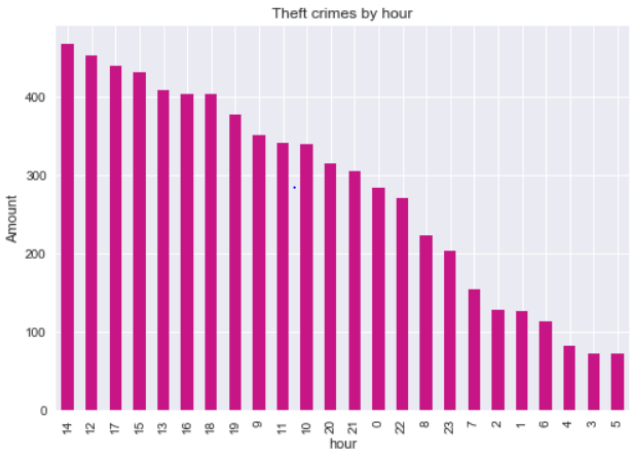
This project visualises the total rate of crime on different days, months and year. It also tells how much crime occurred on that particular day, week and month. So that we can analyse the crime rate in different cities and accordingly take some actions against it. We have also plot different individual crime rate at different time, so that we can analyse a particular crime.

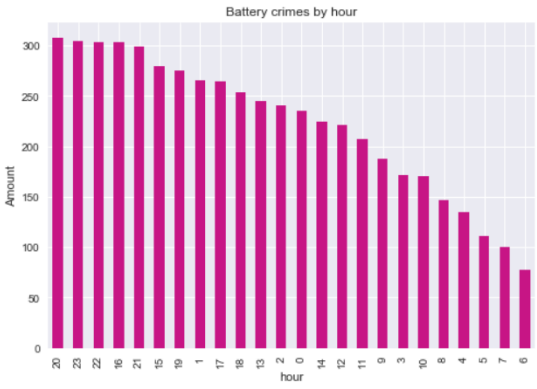
We have also used ggplot to find the exact location(latitude & longitude) of the place of crime and to handle that crime in that area

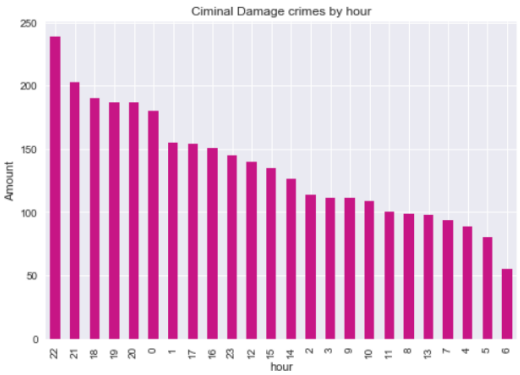


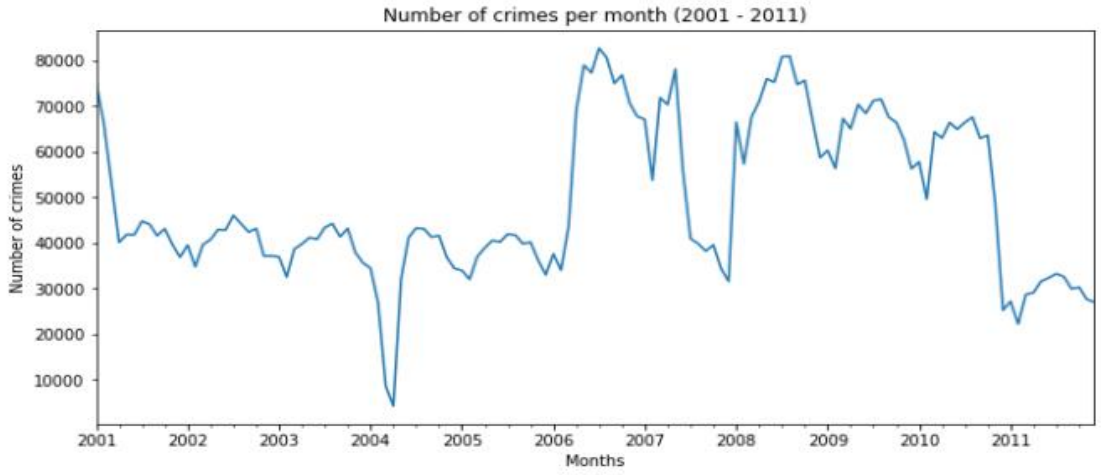


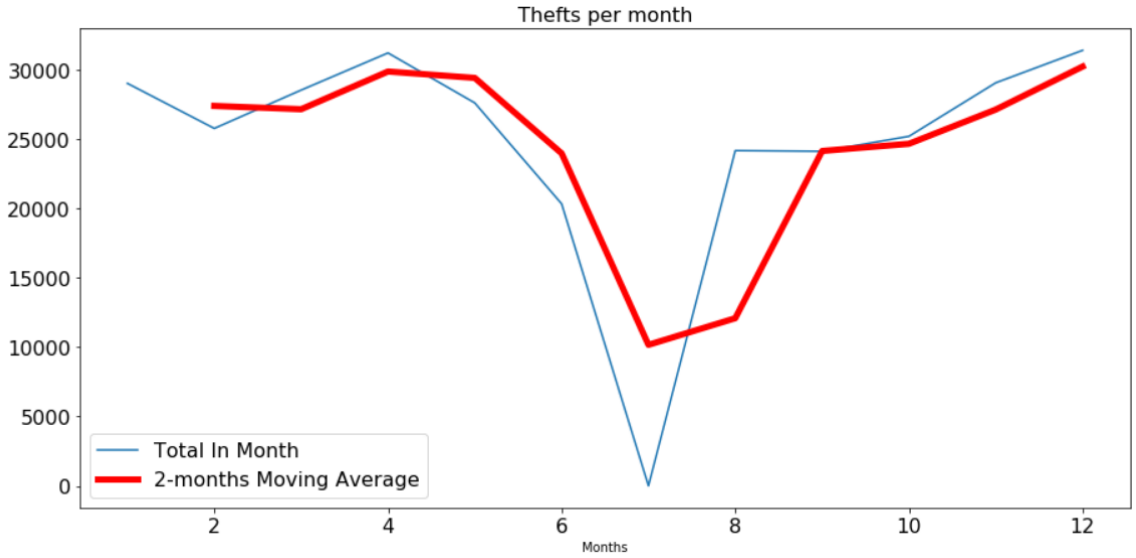





Battery crimes by hour
Amount
hour

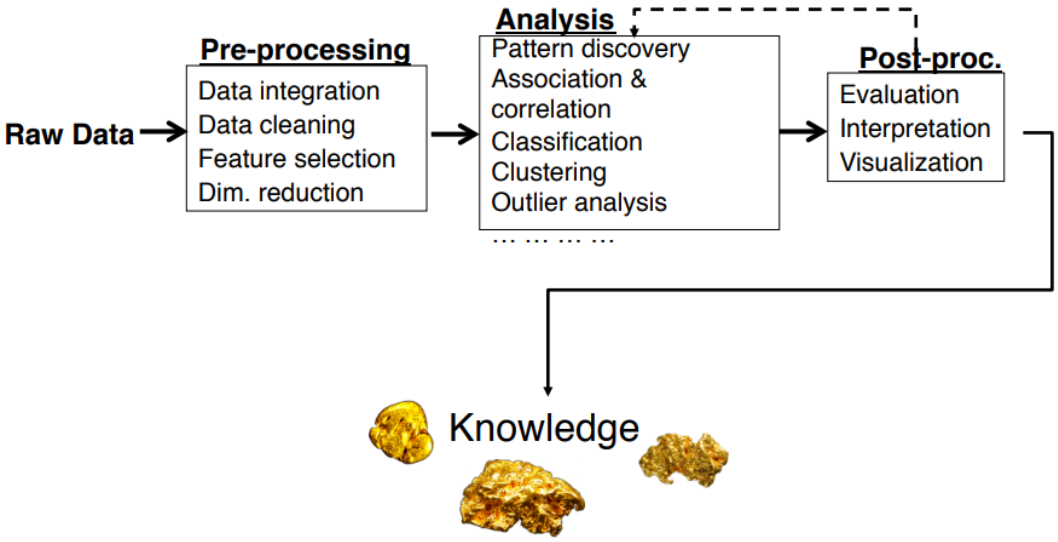

Ciminal Damage crimes by hour
Amount
hour


Number of crimes per month (2001 - 2011)
Number of crimes
Months


Thefts per month
Total In Month
2-months Moving Average
Months

IMPLEMENTATION

FLOWCHART



Pipeline

ALGORITHM

We are using the Classification algorithm in this project to accomplish our task. As there are different types of crimes,we can classify them and perform operation accordingly To accomplish our objective of recognizing crime patterns across the city based on geographical locations, our first measure is dividing the entire city of Chicago into smaller units called cells, each district of Chicago is evaluated as a cell. In the meta-data obtained from the CLEAR system of Chicago Police Department, each criminal record is characterized by several attributes that includes crime description, location, longitudes and latitudes, etc as elaborated in Table 1.

Latitude	The latitude of the location where the incident occurred. This location is shifted from the actual location for partial redaction but falls on the same block.
Longitude	The longitude of the location where the incident occurred. This location is shifted from the actual location for partial redaction but falls on the same block.
Location	The location where the incident occurred in a format that allows for creation of maps and other geographic operations on this data portal. This location is shifted from the actual location for partial redaction but falls on the same block.

Table -1

These attributes comprise the dataset of the system model adopted by our project and will be conducive while plotting the exact locations of the crimes. In addition, the CLEAR system classifies the crimes into 32 different categories as depicted in Table 2.

With all the attributes, we expect to depict the pattern of each crime-type across the City of Chicago for an entire year. For a sound prediction of the occurrence of a crime at any location and any hour of a day, it is required to consider the data that is consistent and out of exemptions. Therefore in order to abstain from false conjectures and guarantee a reliable prediction model, we plan on considering the criminal records of past 6 years to train our algorithm. The derived prediction model is then tested against the records from recent years for validation and determining the accuracy rate of our model.

Theft	Battery
Robbery	Criminal Damage
Deceptive Practice	Narcotics
Domestic Violence	Non-Criminal (Subject Specified)
Assault	Criminal Trespass
Gambling	Arson
Burglary	Prostitution
Concealed Carry License Violation	Human Trafficking
Motor Vehicle Theft	Weapons Violation
Homicide	Offense involving Children
Crime Sexual Assault	Sex Offense
Obscenity	Non-Criminal
Liquor Law Violation	Interference with Public Officer
Kidnapping	Public Peace Violation
Intimidation	Stalking
Public Indecency	Ritualism

Table - 2

COMPLEXITY ANALYSIS

1. Data Exploration

The data was cleaned and preprocessed. The overall dataset includes data between 2001-2017. We need to handle the missing values in the dataset so that it will give us correct result.

2. Data Extraction

There are so many tools for data Preprocessing like Stanford Visualization Group's Data Wrangler [2], Python pandas These tools are fantastic and can save hours. There is overlap in their functionality as well. Using Pandas we find out that based on our goal in the project, the Chicago Crime dataset requires one of the most important data pre-processing procedure which is cleaning. Our data need to be clean by:

Removing duplicate rows

Removing missing values (etc. Null/NA values) in the dataset.

RESULT AND ANALYSIS

In this part, we visualized the data using different ways. We analysed trend of crime occurrence for each year (figure 1). Then, we plotted crime occurrence rates of the following: crime type, scene of crime, hour-day-month of crime (figures 2-6). This gives us a better understanding of the major crimes that occur. We interpret that theft has the highest percentage and is the crime type with the highest crime rate. We also identified locations that are more prone to crimes, street being the scene with highest crime rate.

The crime rate with respect to time of day, day of week and month, depicts which crime happens the most and at what time. Observation says that higher crime intensity is between the duration 12:00PM – 18:00PM. We did breakdown the crimes by location of crime scene, hour of the day, weekday and month as depicted, to see if we can spot any trends.

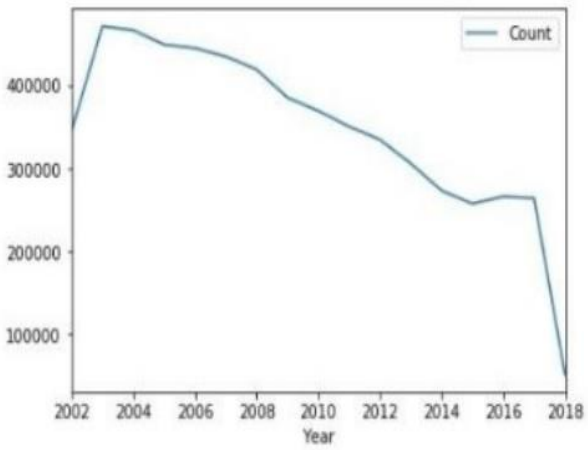


Figure 1: Crime Trend Analysis

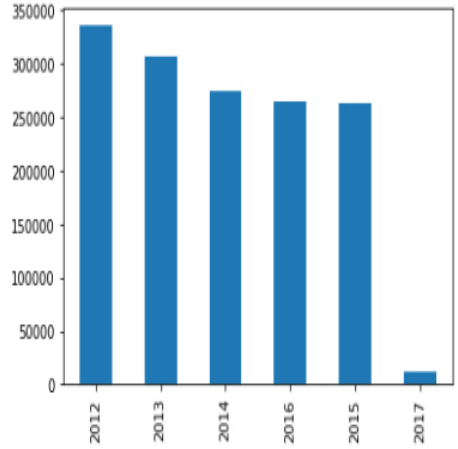


Figure3: Max crime type

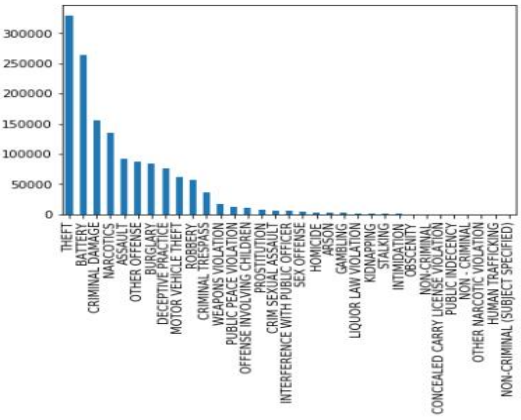


Figure-2: Crime rate V/S Crime type

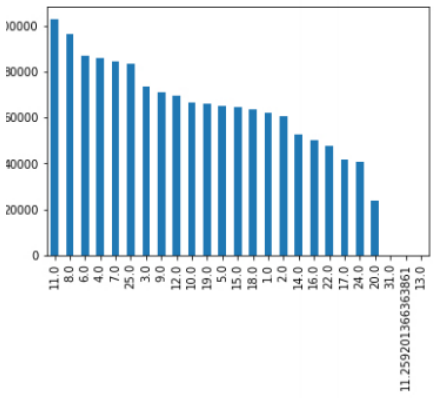


Figure-4: Max crime in districts

Figure-5 Crimes per month

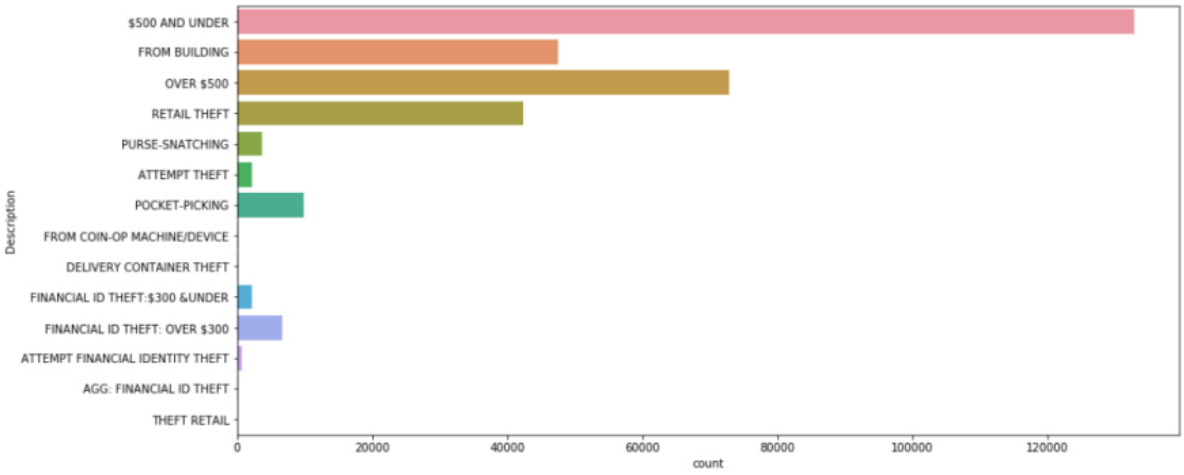


Figure-6 Max crime

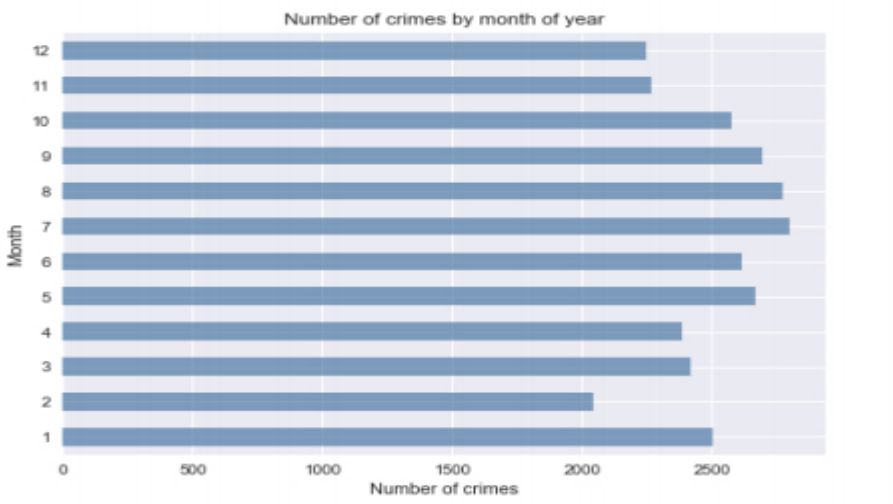


Figure-7 Crime in month

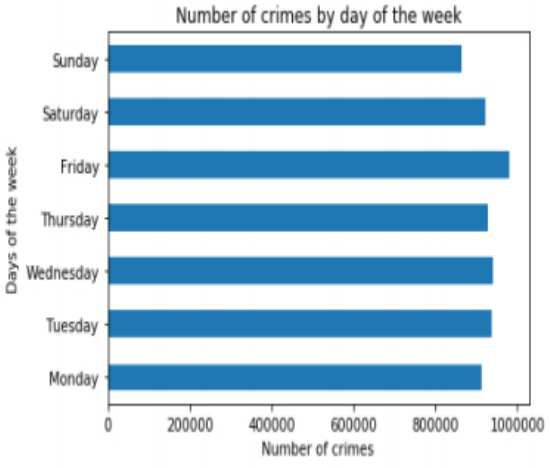


Figure-8 Crime in week days

We want to predict severity of the crime, so we dropped the exact crime type for that point. These are the features that required normalization. Other features for our dataset are all categorical data and have all been converted to dummy variables. We performed split on the 80000 sample records, into training set and test set, and normalized by themselves. Then we performed exploratory data analysis to see what features can be important.

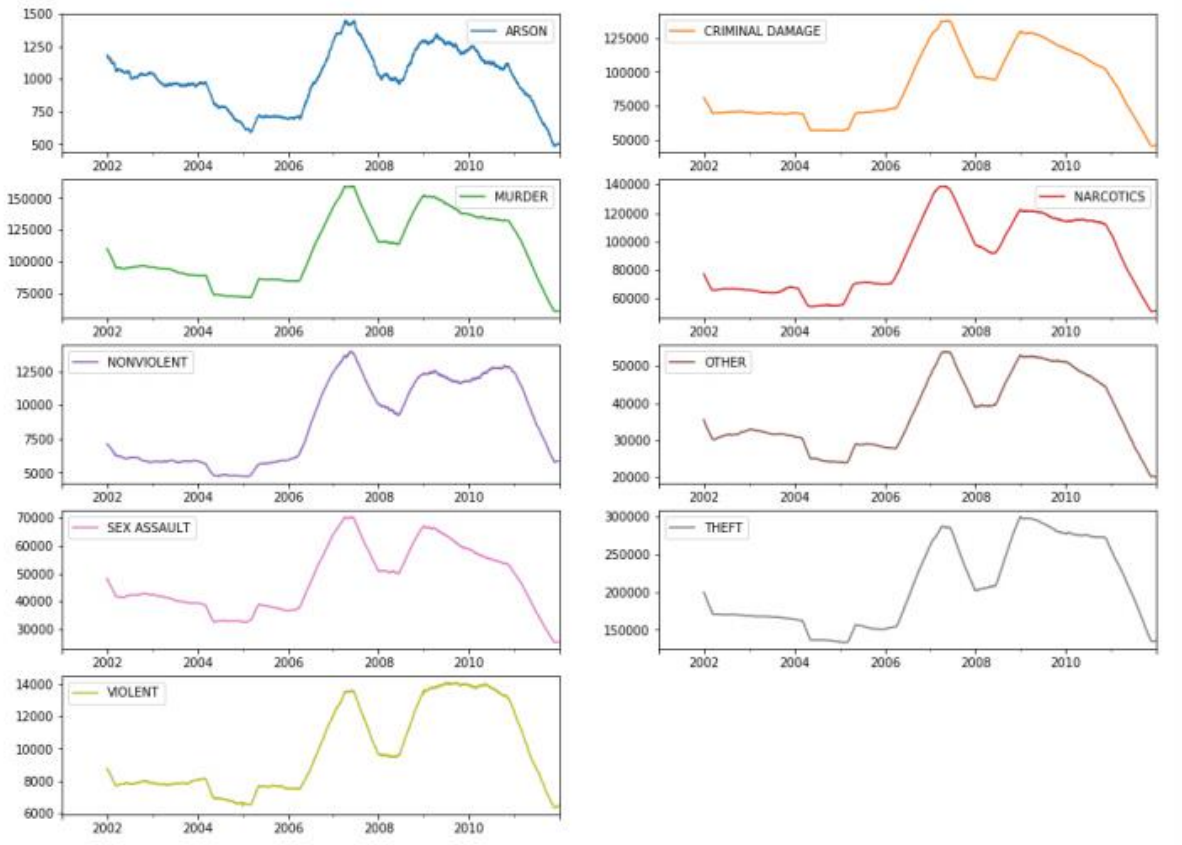


Figure-9 Analysis of different crimes

The density of crime occurrence is demonstrated using a heat map (figure 11). The graph shows the geographical distribution of the crimes in the city of Chicago. The intensity of the color shows the number of crimes. The higher the intensity and higher is the number of crime in that particular area of City of Chicago. This could give a quick insight on which region in the City of Chicago ranks the highest in Crime Rates. The graph is plotted based on the latitude and longitude of the crimes.

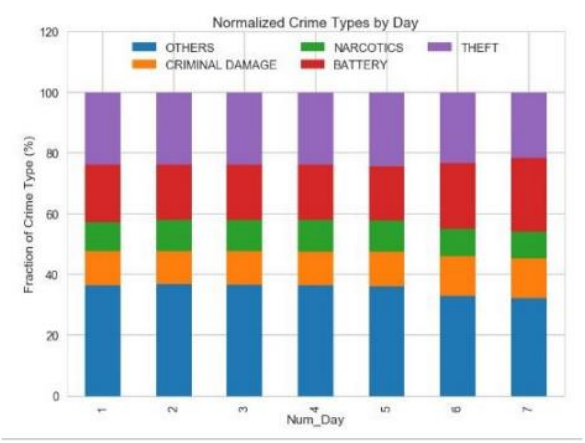


Figure-10 Normalized Crime Types by Day

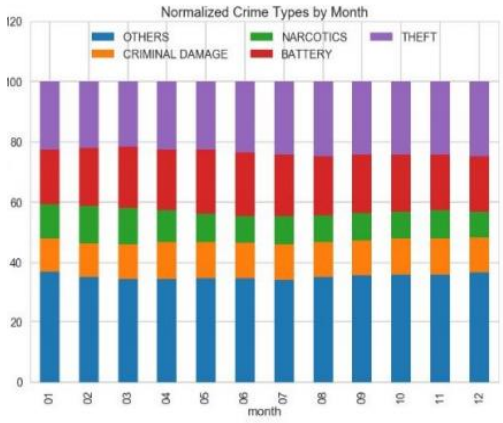


Figure 11: Normalized Crime types Month

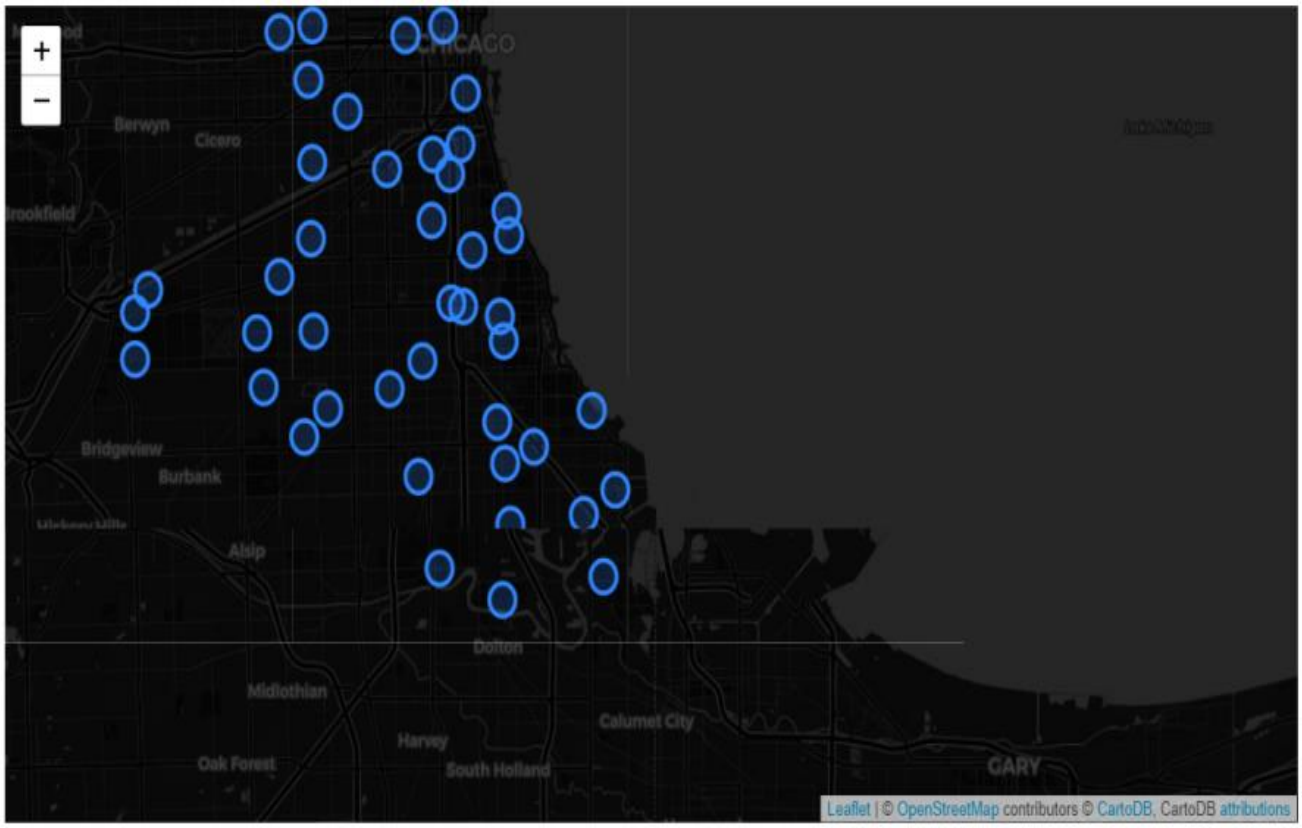


Figure-12 Identification of crime location

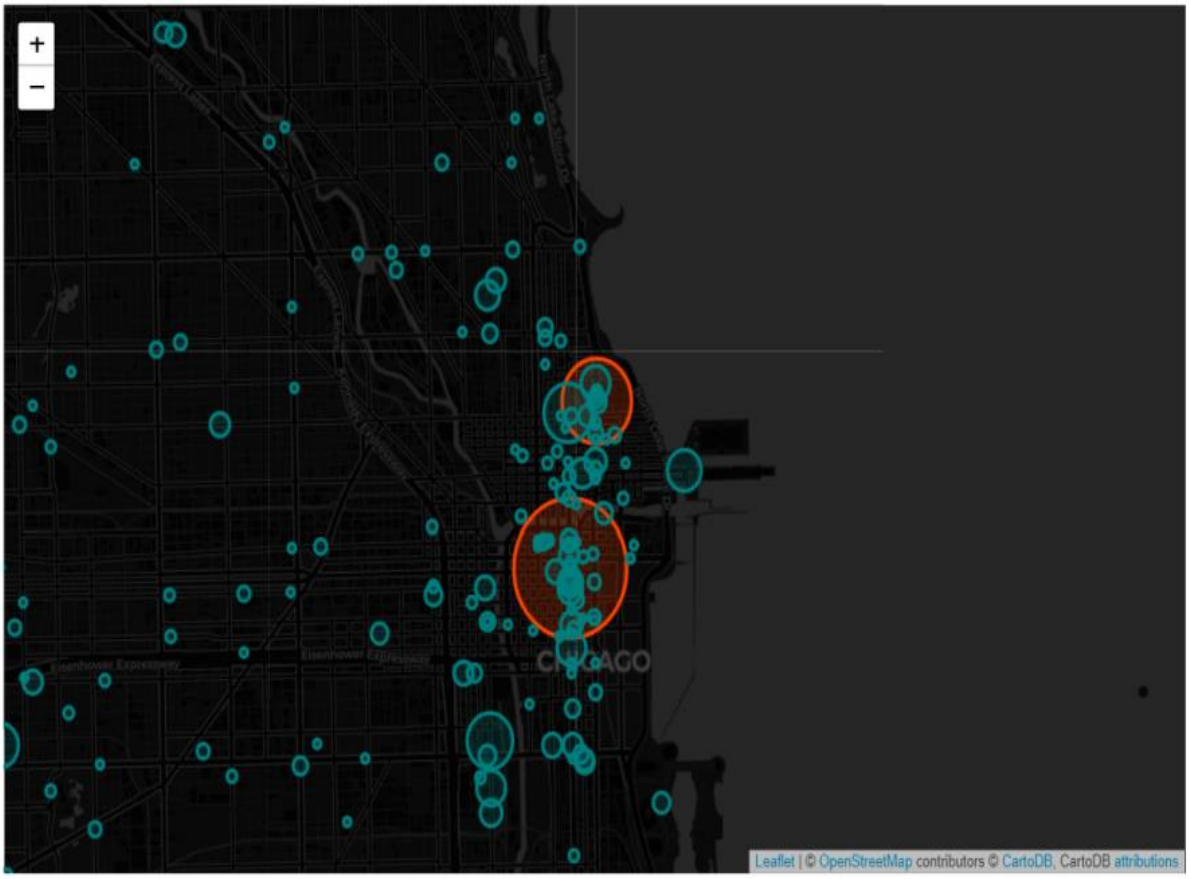


Figure-13 Max crime area

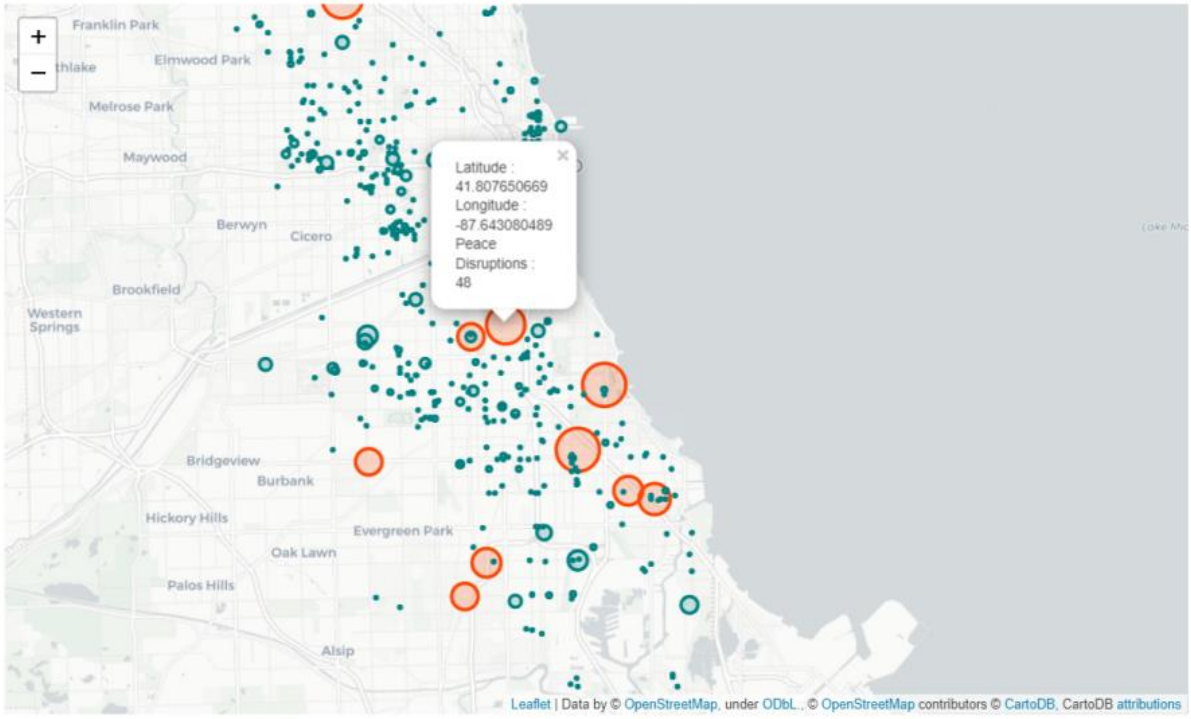


Figure-14 Latitude and longitude

FUTURE WORK

As for future work, in order to boost the classification accuracy, it will be necessary to incorporate other information. For example, the police department may focus on solving a specific type of crime during a specific period of time, which may reduce the occurrence of that type of crime. Additionally, some events and the outcomes of the events may be associated with some crime types, for example basketball games, baseball games and elections. Weather information and classification of buildings can also be incorporated. It will be interesting to see whether these other features can help the classification.

Secondly, the police department can take the safety precautions for the civilians accordingly so that people can live safely and it will be easy for the department to handle all the situations

REFERENCES

[1] Kassen, M. (2013). A promising phenomenon of open data: A case study of the Chicago open data project. *Government Information Quarterly*, *30*(4), 508-513.

[2] Nakaya, T., & Yano, K. (2010). Visualising crime clusters in a space-time cube: An exploratory data-analysis approach using space-time kernel density estimation and scan statistics. *Transactions in GIS*, *14*(3), 223-239

[3] Ratcliffe, J. H. (2000, September). Visualising crime hotspots and making sense of high volume crime. In *a paper presented to the Australian Institute of Criminology conference Crime Mapping: Adding Value to Crime Prevention, Adelaide*.

[4] Shiode, S., Shiode, N., Block, R., & Block, C. R. (2015). Space-time characteristics of micro-scale crime occurrences: an application of a network-based space-time search window technique for crime incidents in Chicago. *International Journal of Geographical Information Science*, *29*(5), 697-719.

[5] Aurisano, J., Kumar, A., Gonzales, A., Reda, K., Leigh, J., Di Eugenio, B., & Johnson, A. (2015). Show me data": Observational study of a conversational interface in visual data exploration. In *IEEE VIS* (Vol. 15, p. 1).

[6] Goldschneider, M. (2010). GEOGRAPHY OF CRIME. REMARKS ON SPATIAL ANALYSES OF CRIME WITH THE USE OF DIGITAL TECHNOLOGIES. *Archiwum Kryminologii*, (32), 23-44.

[7] Rossy, Q., & Ribaux, O. (2014). A collaborative approach for incorporating forensic case data into crime investigation using criminal intelligence analysis and visualisation. *Science & Justice*, *54*(2), 146-153.

[8] Brunsdon, C., Corcoran, J., & Higgs, G. (2007). Visualising space and time in crime patterns: A comparison of methods. *Computers, environment and urban systems*, *31*(1), 52-75.

[9] Gottschalk, P., & Tolloczko, P. C. (2007). Maturity model for mapping crime in law enforcement. *Electronic Government, An International Journal*, *4*(1), 59-67.

[10] Ratcliffe, J. H., & McCullagh, M. J. (1999). Hotbeds of crime and the search for spatial accuracy. *Journal of geographical systems*, *1*(4), 385-398.

[11] Rauscher, J., Swiezinski, L., Riedl, M., & Biemann, C. (2013, June). Exploring cities in crime: significant concordance and co-occurrence in quantitative literary analysis. In *Proceedings of the Workshop on Computational Linguistics for Literature* (pp. 61-71).

[12] Bonatsos, A., Middleton, L., Melas, P., & Sabeur, Z. (2013, October). Crime open data aggregation and management for the design of safer spaces in urban environments. In *International Symposium on Environmental Software Systems* (pp. 311-320). Springer, Berlin, Heidelberg.

[13] Vasiliauskas, D., & Beconytė, G. (2016). Cartography of crime: Portrait of metropolitan Vilnius. *Journal of Maps*, *12*(5), 1236-1241.

[14] Brunsdon, C., & Corcoran, J. (2006). Using circular statistics to analyse time patterns in crime incidence. *Computers, Environment and Urban Systems*, *30*(3), 300-319.

[15] Oatley, G. C., & Ewart, B. W. (2003). Crimes analysis software:'pins in maps', clustering and Bayes net prediction. *Expert Systems with Applications*, *25*(4), 569-588.

[16] Sathyadevan, S., & Gangadharan, S. (2014, August). Crime analysis and prediction using data mining. In *2014 First International Conference on Networks & Soft Computing (ICNSC2014)* (pp. 406-412). IEEE.

[17] Ratcliffe, J. H. (2002). Aoristic signatures and the spatio-temporal analysis of high volume crime patterns. *Journal of quantitative criminology*, *18*(1), 23-43.

[18] Corcoran, J., Wilson, I. D., Lewis, O. M., & Ware, J. A. (2001, October). Data clustering and rule abduction to facilitate crime hot spot prediction. In *International Conference on Computational Intelligence* (pp. 807-821). Springer, Berlin, Heidelberg.

[19] Adderley, R. W., & Musgrove, P. (2001). Police crime recording and investigation systems–A user's view. *Policing: An International Journal of Police Strategies & Management*, *24*(1), 100-114.

[20] Kang, H. W., & Kang, H. B. (2017). Prediction of crime occurrence from multi-modal data using deep learning. *PloS one*, *12*(4), e0176244.