

CRIME STATISTICS ANALYSIS

¹Pallavi N., ²Aiman Khan, ³Anjali Sahu, ⁴Mallika Mondal

¹Assistant Professor, ^{2,3,4}Student

^{1,2,3,4}Computer Science & Engineering

^{1,2,3,4}Atria Institute of Technology, Bengaluru, India

Abstract : Criminal activities all over the world is a challenging issue for most of the government authorities and increased crime rates play a decisive role in hindering the growth of a nation. There is a strong need to monitor and analyze criminal activities across the country as well as regulating them. The best way to overcome this challenge is to first recognize the areas where the types of crime occur and to derive conclusions and suggest measures based on analysis of statistics available.

Index Terms - Crime Analysis, Machine Learning, Prediction, Crime Clustering Analysis

I. INTRODUCTION

All over the world, police agencies use various technologies to tackle the issues related with the crimes occurring over a daily basis. Machine learning tools can be used through different ways to map the occurring and re-occurring of a crime over a particular region. A number of techniques were used for analysis and several maps were prepared in accordance of data present for the mapping of the clusters or the individual mapping of the crime. Though we cannot predict the victims of the crime but we can predict the probability of the place represents the occurrence of that crime.

The results that we obtain after the implementation may not be 100% accurate but it helps in reducing crime rates by providing security in the areas that are much prone to the occurrence of crime. A huge collection of data about the crimes occurring in a particular region is needed that will help us to represent and visualize the data accurately using machine learning tools. The steps involved in crime analysis is depicted below:^[1]

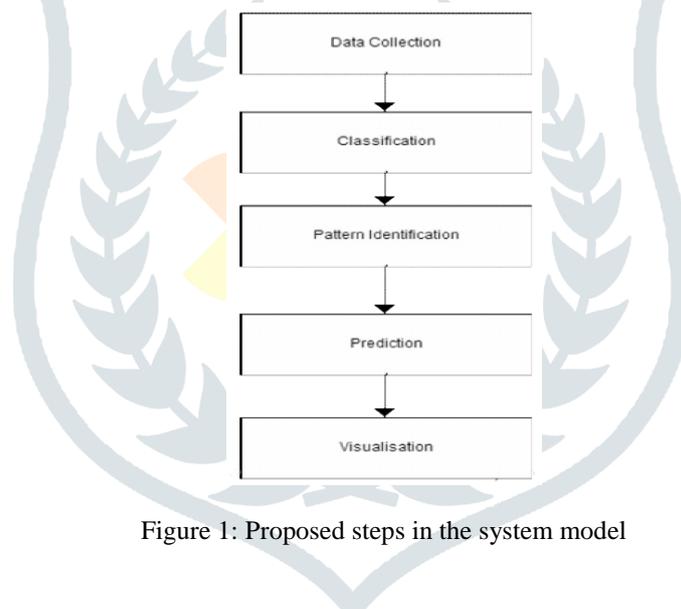


Figure 1: Proposed steps in the system model

II. RELATED WORK

There have been various approaches in this domain, mainly in the analysis area, that deals with basic analysis and mining methods. We can further divide the mining approaches into classification and clustering, for base level understanding of the patterns.

A recent GIS implementation in Chennai ^[2] city was done to map the various property related crimes and to understand the crime pattern analysis of the area studied. Various GIS techniques were used and several maps were created to show the patterns analyzed from the datasets collected. Another theoretical model based on data mining techniques such as classification and clustering ^[3] was used extensively in terms of analysis and discovery of patters from the crime datasets of England.

‘Predictive Crime Mapping’, according to Fitterer et al was to spatially predict commercial and residential break and entries (B&Es) in Vancouver, Canada. ^[4] The results of the analysis done found patterns in the characteristics of BNEs, including progressive decrease in annual trends in the frequency of B&Es. Another study based on crimes related to women was implemented in India that helps provide preventive strategies through visualization and prediction. ^[5]

III. PROPOSED APPROACH

An understanding of where and why crimes occur can help to efficiently fight crime. Through the use of machine learning libraries, maps provide a swift access to regions with higher concentrations of crime, helping authorities with proper action. Simpler maps can help the police with patrol activities, whereas more complex maps showcase trends in criminal activity throughout the region. ^[6]

The approach described in this paper involves data-preprocessing and analysis on the process data for representation and visualization. The major purpose of data pre-processing is to process the existing data so as to produce a working dataset on which analysis can be done. Before performing any sort of analysis, understanding the data at hand is of importance. This stage in itself involves a number of processes, that include –

3.1 Data Collection

This process involves acquiring and organizing the data relevant to the problem. For this project there were two different data sets: Vancouver Crime 2002 -2018 from the Vancouver data catalogue at: data.vancouver.ca. Weather data from 2013 - 2017 collected at YVR INT A weather terminal. The original dataset of the various crime statistics can be seen in Fig. 2. Storage of data in an organized manner makes accessing and performing calculations over it faster, as the entire data can be traversed at the same time while analyzing.

	TYPE	YEAR	MONTH	DAY	HOUR	MINUTE	HUNDRED_BLOCK	NEIGHBOURHOOD	X	Y
0	Break and Enter Residential/Other	2003	12	9	15.0	45.0	1X E 57TH AVE	Sunset	492299.45	5451762.38
1	Theft of Vehicle	2003	1	11	22.0	0.0	55XX ARGYLE ST	Kensington-Cedar Cottage	494802.88	5453540.96
2	Theft from Vehicle	2003	1	17	21.0	48.0	16XX DAVIE ST	West End	489906.41	5459175.15
3	Theft of Vehicle	2003	12	5	17.0	0.0	47XX CULLODEN ST	Kensington-Cedar Cottage	494258.53	5454405.92
4	Mischief	2003	9	12	3.0	30.0	31XX E 51ST AVE	Killarney	497055.35	5452244.07
5	Theft of Vehicle	2003	11	23	21.0	0.0	47XX DUMFRIES ST	Kensington-Cedar Cottage	494639.97	5454370.24
6	Theft from Vehicle	2003	12	2	17.0	45.0	5XX MAIN ST	Central Business District	492741.41	5458535.19

Figure 2: Original dataset after pre-processing and adding additional features

3.2 Pattern Identification

With the data having been classified based on existing relationships between the members, the next process involves identifying interesting patterns and trends that exist over the entire dataset, as shown in Fig 3. Finding the common patterns helps with visualizing regions of common activity as well as help with efficient analysis and proper results.

```

Break and Enter Commercial          34928
Break and Enter Residential/Other    62030
Homicide                             226
Mischief                             73221
Offence Against a Person            55726
Other Theft                          54807
Theft from Vehicle                   179700
Theft of Bicycle                     26942
Theft of Vehicle                     39164
Vehicle Collision or Pedestrian Struck (with Fatality)  263
Vehicle Collision or Pedestrian Struck (with Injury)    22756
Name: TYPE, dtype: int64

```

Figure 3: Different types of crime identified from the dataset

3.3 Prediction

Based on common trends over the entire dataset, predicting hotspots of major activity occurring over that particular region of crime and type of crime can be predicted. The predicted results shows that crime rates can be identified on an hourly basis based on neighborhood, as shown in Fig 4. Various algorithms are used for predictive analysis such as Regression, Naïve Bayes, Classification, etc. that can be used to improve accuracy of analysis based on the dataset. Out of these, unsupervised K-Means algorithm is successfully used to showcase the clusters of crimes per hour on the basis of neighbourhood of occurrence. ^[7]

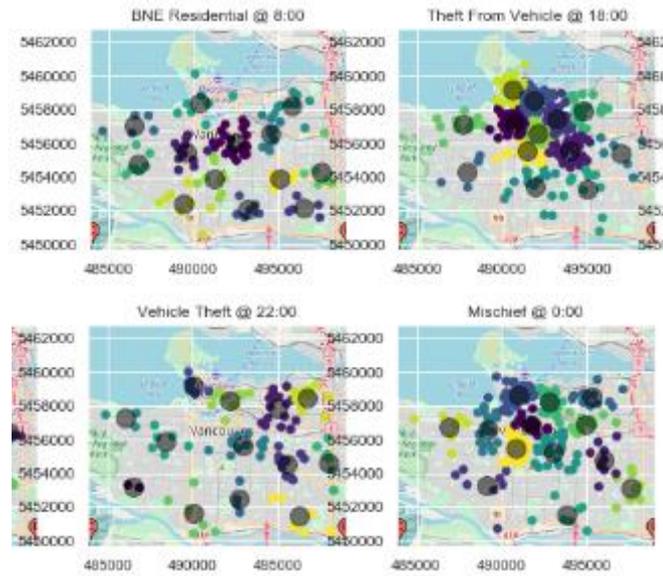


Figure 4: Clustering of crime types based on hourly data per neighbourhood

3.4 Data Visualization

Visualizing the data helps to sort out patterns in a much easier way. [8] Data, in general, are visualized in the form of graphs or histograms – which are a representation of numerical data. As a result of predictive analysis, the regions of interest and patterns are visualized with the use of maps. Patterns, trends and correlations that might go undetected in text-based data can be exposed and recognized easier with various data visualization approaches.

3.4.1 Graph Visualization

This type of visualization displays information as a series of data points called “markers” which are connected by straight lines. This type of plot is used to visualize the crime trends in the dataset analysed over intervals of time - a time series, as shown in Fig. 5.

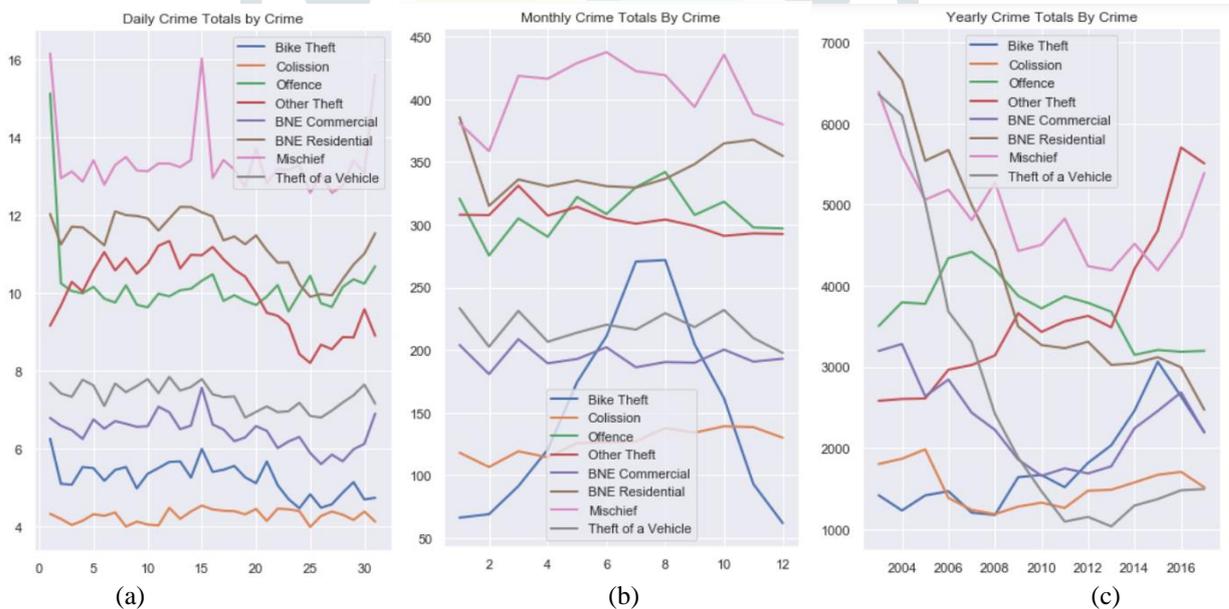


Figure 5: Various graph visualizations: (a) Yearly crime totals by crime (b) Monthly crime totals by crime (c) Daily Crime totals by crime

3.4.2 Map Visualization:

This type of visualization projects the results of the pre-processed data in the form of a geographic map, that shows the various types of crimes and their regions of occurrence, as shown in Fig. 6.

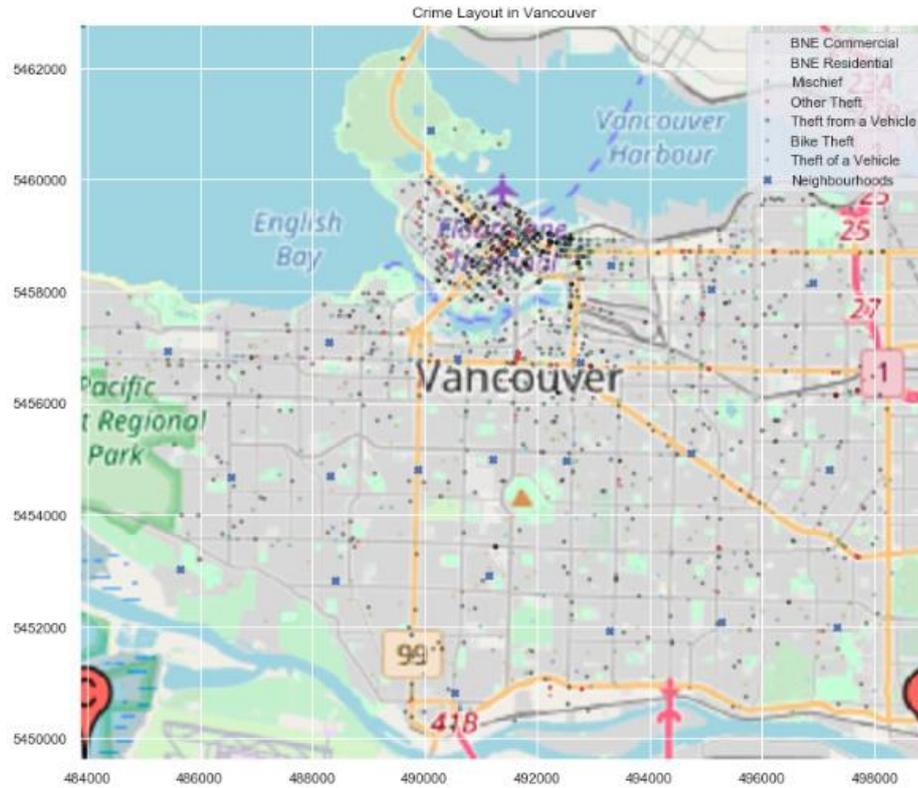


Figure 6: Map of Vancouver showing the various types of crimes and their region of occurrence

3.4.3 Heatmap Visualization

A heatmap is a two-dimensional graphical representation of data where colors are used to represent the average number of crimes per day and month extracted from the dataset. These colors also give insight on intensities, as shown in Fig. 7.

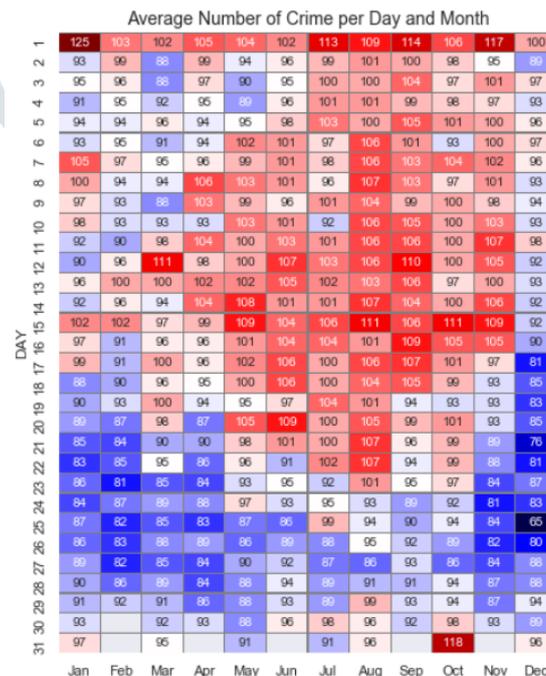


Figure 7: Heatmap of average number of crime per day and month

IV. CHALLENGES FACED

The main challenges faced are:

1. The crime rates do not remain constant in any region or country. This brings up a huge challenge of applying the approach in the same way for different amounts of data. This may affect the visualization.^[9]
2. Sometimes the data that we collect from different sources for the implementation is incomplete or insufficient. This type of data can lead to wrong prediction of crime rate, that will in turn lead to lack of proper preventive measures in the region being investigated.
3. Law Enforcement Departments play a very important role in providing us the statistics related with crime. This forms one of the biggest challenge when collecting data for implementation.
4. The more accurate is the data, the more accurate is the program in providing the output. The accuracy of the whole program depends on the accuracy of the training set.^[10]

V. FUTURE ASPECTS

Although the proposed approach provides an efficient method towards visualizing, crime mapping and analysis of criminal activity patterns in an area based on the statistics observed during the past years, there are some aspects that can be considered for development in the future such as:

1. To forecast criminal activity and other related phenomena in the form of citizen-friendly applications available for different time and space scales.
2. To evaluate the effects of actions and measures undertaken which effectively help in regulating crime and it's socio-economic effects.^[11]
3. For regular updating of database in the application to make them more aware of the region with high criminal activity for cautionary measure.
4. To work on larger datasets than the current database, and to keep it's correctness estimation similar or better to the current approach.
5. To allow for co-operation with governmental authorities as well as co-operation of civilians for future preventive and regulatory measures.

VI. CONCLUSION

Although the active potential of the crime rate in accordance of solving of the crime is still not evitable, this paper makes an effort to explain the way how crimes and crime is concentrated in a region and how it helps government authorities to plan a secure region in order to avoid any more crimes in that particular region.

REFERENCES

- [1] Shiju Sathyadevan, Devan M.S, Surya Gangadharan. S, Crime Analysis and Prediction Using Data Mining, DOI: 10.1109/CNSC.2014.6906719
- [2] K. Jaishankar, S. Shanmugapriya, V. Balamurugan, Crime Mapping in India: A GIS Implementation in Chennai City Policing, The International Association of Chinese Professionals in Geographic Information Sciences, 1082-4006/04/1001/1001-20
- [3] Rasoul Kiani, Siamak Mahdavi, Amin Keshavarzi, Analysis and Prediction of Crimes by Clustering and Classification, (IJARAI) International Journal of Advanced Research in Artificial Intelligence, Vol. 4, No.8, 2015
- [4] J. Fitterer, T.A. Nelson, F. Nathoo, Predictive Crime Mapping, Police Practice and Research: An International Journal, DOI: 10.1080/15614263.2014.972618
- [5] Tushar Sonawanev, Shirin Shaikh, Shaista Shaikh, Rahul Shinde, Asif Sayyad, Crime Pattern Analysis, Visualization And Prediction Using Data Mining, IJARIE-ISSN(O)-2395-4396, Vol-1 Issue-4 2015
- [6] Peter Harrington, Machine Learning in Action, MANNING, Shelter Island, 2012, www.library02.embl.de
- [7] David E. Goldberg, John H. Holland, Genetic Algorithms and Machine Learning, October 1988, Volume 3, Issue 2–3, pp 95–99
- [8] Usama Fayyad, Georges G. Grinstein, Andreas Wierse, Information Visualization in Data Mining and Knowledge Discovery, Morgan Kauffman Publishers, ISBN 1-55860-689-0
- [9] Chong Ho Yu, Exploratory Data Analysis, DOI: 10.1093/OBO/9780199828340-0200
- [10] Tom M. Mitchell, The Discipline of Machine Learning, July 2006, CMU-ML-06-108
- [11] Gregory F. Cooper, Constantin F. Aliferis, Richard Ambrosino, John Aronis, Bruce G. Buchanan, Richard Caruana, Michael J. Fine, Clark Glymour, Geoffrey Gordon, Barbara H. Hanusa, Janine E. Janosky, Christopher Meeke, Tom Mitchell, Thomas Richardson, Peter Spirtes, An evaluation of machine-learning methods for predicting pneumonia mortality, [https://doi.org/10.1016/S0933-3657\(96\)00367-3](https://doi.org/10.1016/S0933-3657(96)00367-3)