City Crime Profiling Using Cluster Analysis

Priyanka Gera[1], Rajan Vohra[2]

[1]Student of M.Tech(Computer Science),
[2]Head of Department(Computer Science & Information Technology)
PDM College of Engineering ,Sector 3A, Sarai Aurangabad,
Bahadurgarh, Haryana, India

Abstract - Crimes cause terror and cost our society dearly in several ways. Data mining can be used to model crime profiling. Here we look at use of clustering algorithm for a data mining approach to analyze the crimes patterns. We will look at k-means clustering to aid in the process of crime profiling. We applied these techniques to primary crime data from Delhi police first information report (FIR) records. In this paper, firstly we desire to estimate which type of crime is dominant in Delhi city, India. Accordingly crime is divided into three types heinous crime, non-heinous crime and special & local laws violation. Second estimation is to find which area categories are more sensitive towards, areas categories which are considered are slums, residential, commercial, VIP zones, travel points and markets. Third is to show distributions of each crime type in every area category.

Keywords -Data mining, Crime profiling, clustering, kmeans, weka.

I. INTRODUCTION

Historically solving crimes has been the prerogative of the criminal justice and law enforcement specialists. With the increasing use of the computerized systems to track crimes, computer data analysts have started helping the law enforcement officers and detectives to speed up the process of solving crimes. Here we will take an interdisciplinary approach between computer science and criminal justice to develop a data mining paradigm that can help solve crimes faster. More specifically, we will use clustering based models to help in identification of crime profiles.

In ordinary language, the term crime denotes an unlawful act punishable by the state. Crime is referred to as a comprehensive concept that is defined in both legal and non-legal sense (Akpinar et al, 2005.

Various agencies such as Delhi police, National Crime Records Bureau (NCRB) [1], Delhi, India, District Crime Records Bureau (DCRB), India are working rigorously to combat the crime. But the challenges to analyze the crime profiles and policing strategies is becoming more difficult as the crime rate is increasing day by day

The usage of data mining concept help to explore the enormous data and making it possible in reaching the ultimate goal of crime profiling .This study attempted to extract crime profiles hence police officials can make use of these profiles in their day-to-day battle against crime. Police officials in the crime prevention and investigation authority of the respective region can make use of the results of this study in order to make optimal deployment of resource in crime prevention. Moreover, the output of the study shall be used for designing appropriate training

programs and crime prevention and investigation strategies. The output of the study shall also be used as a benchmark for police officials as well as a source of methodological approach for studies dealing on the application of data mining on crime management as well as other similar areas.

II. RESEARCH METHODOLOGY

For solving the problem some research techniques and methodologies are used for obtaining the desired result. Main steps under the research methodologies are:-

- ➢ *Review of literature* – Relevant literatures on data mining techniques and crime were reviewed. The potential of data mining in general and particularly successful data mining applications in crime prevention were assessed.

- ➢ *System study* - In order to define problem properly, primary information regarding crime in Delhi was collected by interviewing concerned officers in the Delhi police commission as well as through National Crime Records Bureau (NCRB), Delhi, India statistics and Delhi police website. The research strategy adopted was to conduct a survey to gather information required to solve the problems specified in the work.

- ➢ *Define scope of work* - Then based on the information obtained from system study determination of nature and definition of problem.

- ➢ *Study concepts needed* – One of the critical tasks that have to be performed at this step is selection of software that supports the data mining techniques that are to be employed in the study. In conducting this research weka software are employed for reasons of accessibility and familiarity. The concepts required for work like clustering, k-means clustering were studied.

- ➢ *Finalizing database attributes*– Attributes and structure of the database was thoroughly studied for finding out useful attributes from the National Crime Records Bureau (NCRB) [1], Delhi, India statistics and Delhi police website [2].

- ➢ *Primary data collection*: In order to define the research problem properly, primary data was collected by interviewing concerned officers in the Delhi police commission as well as through National Crime Records Bureau (NCRB), Delhi, India statistics and through semi-structured interviews, group discussion, participant observation, documentation analysis and questionnaires. After the data are collected, orders such as processing and cleansing are imposed in order to make the data more suitable for the particular data

mining software, which are used in the study.

➢ *Data Population and analysis* - Database is then populated and data analysis is done using tool.

➢ *Results of data analysis* – Tabulated results of data analysis is presented after which interpretation of results is done and explained in detail.

III. RESEARCH BACKGROUND

Data mining is a powerful tool which enables investigators to explore large criminal and crime databases quickly and efficiently.

D.E. Brown (1998) [3] constructed a software framework called ReCAP (Regional Crime Analysis Program) for mining data in order to catch professional criminals using data mining and data fusion techniques.

In 2009, Li Ding et al. [4] propose an integrated system called PerpSearch that takes a given description of a crime, including its location, type, and the physical description of suspects (personal characteristics or vehicles) as input. To detect suspects, the system will process these inputs through four integrated components: geographic profiling, social network analysis, crime profiles, and physical matching. Essentially, geographic profiling determines "where" the suspects are, while other components determine "who" the suspects are.

In 2012, K. Zakir Hussain, et al. [5] presented Criminal behavior analysis by using data mining techniques. He elaborated the criminal behavior analysis of the offenders by using data mining techniques.

In 2013,Kaumalee Bogahawatte & Shalinda Adikari [6] developed a system named Intelligent Criminal Identification System (ICIS) that can potentially identify a criminal based on the evidences collected from the crime spot. The solution has provided for three crime categories namely robbery, burglary and theft out of 2 1 categories of grave crimes.

IV. CONCEPTUAL FRAMEWORK

This paper is further categorized in three sub parts which are as follows:

A. *Tier 1: Crime profiling in city based on types of crimes*

It is crime profiling based on types of crimes. The result will show three clusters showing crime type in city.

Crime types considered are:

1) Heinous crimes: It includes crimes like murder, attempt to murder, robbery, roit, rape, dacoity,etc.

2) Non-heinous crimes: Crimes considered under non-heinous crimes are Forgery, Counterfeiting, Burglary, Fraud, Theft, Pick pocketing, M.O women, Kidnapping , Arson, etc.

3) Special and local laws: Special and local laws includes violation of some special and local laws such as Arms act, Narcotics drugs and psychotropic substance act, Gambling act, Exercise act, Prohibition act, Child labour, Child punishment, etc.

Primary database is created in accordance of First Information Report (FIR) records of Delhi police, India [7]. Then random sampling is done to refine database. Crimes

in database records are categorized in above mentioned crime types based on acts and sections in database. Then data is feed to weka tool and clustering is done.

B. *Tier 2: Crime profiling in city based on types of areas*

It includes crime profiling based on categories of areas. The result will show six clusters showing crime in six types of areas.

Area categories considered for profiling are:

➢ Slums
➢ Residential areas
➢ Commercial areas
➢ VIP zones- high security zone areas for Very important people
➢ Travel points
➢ Markets

Primary database is created in accordance of FIR records of Delhi police, random sampling is done and then database is refined on basis of area category. Then data is feed to weka tool and clustering is done.

C. *Tier 3: Association between crime type and area type*

It is to identify association between crime type and area category . This shows in which area category which type of crime occurs the most and least. For eg: In slums heinous crime occurs x%, non henious y % and specail and loacl laws z%. Database is refined and weka tool is used to perform clustering.
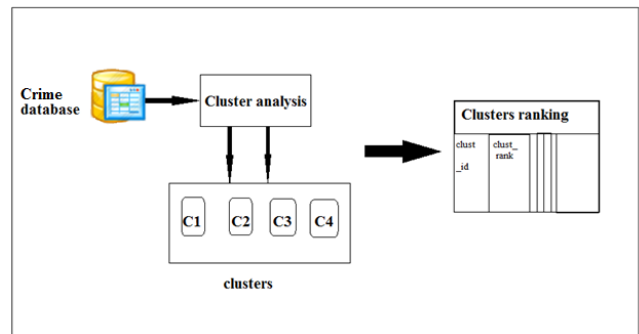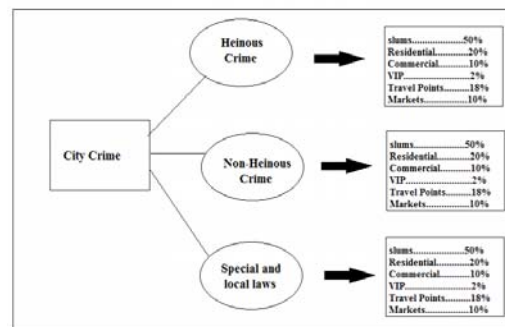

Fig. 1 Crime profiling in city

Fig. 2 Overview of crime profiling

V. RESULTS AND DISCUSSION

A database of 998 records is created based on primary data obtained from various FIR records of different areas of Delhi. Database attributes are FIR number, district where crime incident occurred, police station where report is written, acts and sections which will tell about name of crime occurred, type of information whether it is written or

oral , address where crime occurred which will tell us category of area whether it is slum, residential, commercial, VIP, travel point or market. It also includes crime, crime category, area category and area code.

A. Tier 1: Profiling based on type of crime

Clustering is done based on crime type i.e. heinous crime, non-heinous crime and special & local laws. Here, only one attribute is taken i.e. crime type is taken to find out crime type clusters from whole dataset.
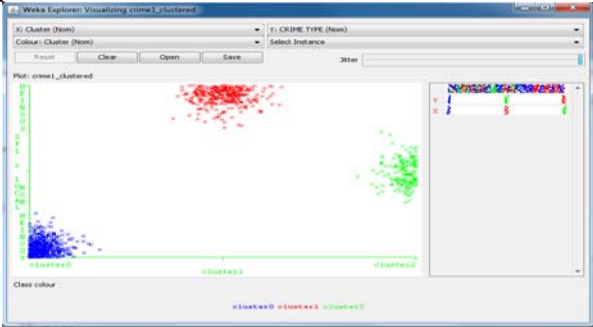

Fig. 3 Graphical representation of crime type clusters

Results of tier 1 are summarized in table below:

Cluster Number	Crime type	No. of records	% of total city crime	Cluster rank
Cluster 0	Non-heinous	492	49%	1
Cluster 1	Heinous Crimes	338	34%	2
Cluster 2	Special and local laws	168	17%	3

Table I Crime profiling based on crime type clusters results

Table I interprets that three clusters are formed with maximum entries in cluster 0, that is, non heinous crime type which is given rank 1. Similarly based on number of entries or percentage of crime type ranks are assigned to clusters.

Thus from table we can conclude that non heinous crime type occurs maximum in city with 49% of total crime, second is heinous crime with 34% of total crime and last is special and local laws with 17% of total crime.

B. Tier 2: Crime profiling based on type of area

Clustering is done based on area category i.e. slums, residential, commercial, VIP zones, travel points and markets. Here, only one attribute is taken i.e. area type is taken to find out area type clusters from whole dataset.
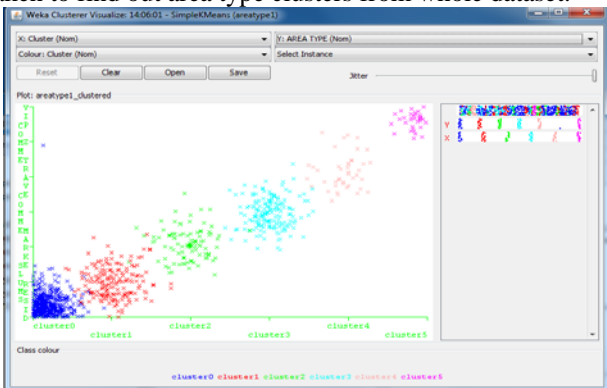

Fig. 4 Graphical representation of area type clusters

Results of tier 2 are summarized in table below:

Cluster Number	Area type	No. of records	% of total city crime	Cluster rank
Cluster 0	Residential	473	47%	1
Cluster 1	Slums	179	18%	2
Cluster 2	Markets	98	10%	4
Cluster 3	Commercial area	172	17%	3
Cluster 4	Travel points	43	4%	5
Cluster 5	VIP Areas	33	3%	6

Table II Crime Profiling Based on Area Type Clusters Results

Table II explains the result of tier 2. In tier 2 six clusters are formed based on area types with clusters 0, 1, 2, 3, 4, 5 for residential, slums, markets, commercial area, travel points, VIP areas respectively. Rank to each cluster is assigned based on percentage of total crime occurring in that area. According to results maximum crime that is 47% of total crime occurs in residential area. Next is slums with 18%, then commercial area, markets, travel points and VIP areas. From these results we can conclude that maximum crime in Delhi city occurs in Residential area and least in VIP areas.

C. Tier 3: Profiling based on particular type crime and particular type of area.

Here, two attributes i.e. crime type and area type are taken from whole dataset.
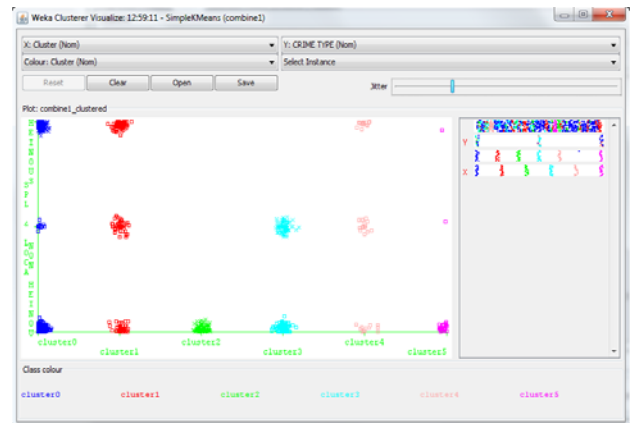

Fig. 5 Graphical representation of clusters with area and crime type.

In Fig. 5 x-axis shows area type and y-axis displays crime type.

Results of tier 3 are summarized in table below:

Crime Type	Residential	Slums	Markets	Commercial	Travel point	VIP zones
Heinous	23.8%	8.3%	0%	0%	1.6%	0.1%
Non heinous	22.1%	4.9%	9.8%	8.01%	1.3%	3.1%
Spcl & Local	1.4%	4.7%	0%	9.2%	1.4%	0.1%

Table III Association between crime type and area type

Table III explains the association between area type and crime type. In residential area heinous crime occurs 23.8%, non heinous 21.1% and special and local is 1.4%. In slum area heinous crime occurs 8.3%, non heinous 4.9% and special and local is 4.7%. In markets heinous crime is nil, non heinous 8.01% and special and local is 9.8%. In commercial area heinous crime does not occurs, non heinous 8.01% and special and local is 9.2%. At travel points heinous crime occurs 1.6%, non heinous 1.3% and special and local is 1.4%. VIP zones heinous crime occurs 0.1%, non heinous 3.1% and special and local is 0.1%.

VI. CONCLUSION AND FUTURE SCOPE

The purpose of this research was to explore the applicability of data mining techniques in the process of crime detection. This paper provides information about crime that is how much percentage of which crime type occurs in Delhi city and which area types are more sensitive to crime and percentage of each particular type of crime occurs in which area. Area types considered in work are: slums, residential,commercial, VIP zones, travel points and markets and crime types considered are: heinous crimes, non-heinous crimes and special and local laws. This provides information that which areas are sensitive towards crime and also association between areas and crime types. This can help Delhi police in analyzing crime profiles. The paper attempted to extract crime profiles hence police officials can make use of these profiles in their day-to-day battle against crime.

There is always remains scope of improvement in any research and so is with this research also. Future scopes can be:

- ➤ Visual and intuitive criminal and intelligent investigation techniques can be developed for crime pattern.
- ➤ Social link networks can be developed to link criminals, suspects, gangs and study their interrelationships.
- ➤ District wise crime mapping can be done.
- ➤ Data warehouse development for tracking crime trends and designing policing strategies.

REFERENCES

[1] http://ncrb.nic.in/
[2] www.delhipolice.nic.in
[3] D.E. Brown, *The regional crime analysis program (RECAP): A frame work for mining data to catch criminals*, in Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics, Vol. 3, pp. 2848-2853, 1998.
[4] Li Ding et al., *PerpSearch: an integrated crime detection system* , 2009 IEEE 161-163 ISI 2009, June 8-11, 2009, Richardson, TX, USA.
[5] K. Zakhir Hussain, M. Durairaj and G. Rabia lahani Farzana ,*Criminal behavior analysis by using data mining techniques*, IEEE-International Conference On Advances In Engineering, Science And Management (ICAESM -2012) March 30, 31, 2012, pp.656-658,2002.
[6] Kaumalee Bogahawatte and Shalinda Adikari, *Intelligent criminal identification system*, The 8th International Conference onComputer Science & Education (ICCSE 2013) April 26-28, 2013. Colombo, Sri Lanka. pp.633-638, April 2013.
[7] http://delhipolice.serverpeople.com/firwebtemp/Index.aspx
[8] P. Thongtae and S. Srisuk, *An analysis of data mining applications in crime domain*, IEEE 8th International Conference on Computer and Information Technology Workshops, (122-126),2008.
[9] Gorr, W. L., Olligschlaeger, A. M., & Thompson, Y, *Short-term time series forecasting of crime.* International Journal of Forecasting, 19, (579-594),2003.
[10] S.Yamuna, N.Sudha Bhuvaneswari ,*Data mining Techniques to Analyze and Predict Crimes*, The International Journal of Engineering And Science (IJES) ,Volume 1 ,Issue 2, (243-247),2012.
[11] S.V. Nath, *Crime pattern detection using data mining*, in Proceedings of the 2006 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology, pp. 41-44,2006.
[12] Chung-Hsien Yu, Max W. Ward1, Melissa Morabito2, and Wei Ding, *Crime forecasting using data mining techniques*, 11th IEEE International Conference on Data Mining Workshops, pp.779-786,2011.
[13] Malathi. A, Dr. S. Santhosh Baboo and Anbarasi. A ,*An intelligent analysis of a city crime cata using data mining*, 2011 International Conference on Information and Electronics Engineering IPCSIT vol.6 (2011) © (2011) IACSIT Press, Singapore, pp.130-134,2011.
[14] Peng Chen, Hongyong Yuan, Xueming Shu,"*Forecasting crime using the ARIMA model,* Fifth International Conference on Fuzzy Systems and Knowledge Discovery 2008 (627-630),2008.
[15] J. Han and M. Kamber, *Data Mining: concepts and techniques*,Morgan Kaufmann publications, pp. 1-39, 2006.
[16] J. Mena, *Investigative Data Mining for Security and Criminal Detection*, Butterworth Heinemann Press, pp. 15-16, 2003.