

# Analysis of Criminal Behavior through Clustering Approach

<sup>1\*</sup>Romika Yadav, <sup>2</sup>Savita Kumari Sheoran

<sup>1,2</sup>Dept. of Computer Science & Engineering Indira Gandhi University Meerpur, Rewari - INDIA

Available online at: [www.ijcseonline.org](http://www.ijcseonline.org)

Accepted: 21/Nov/2018, Published: 30/Nov/2018

**Abstract** – The spatio-temporal modeling of a social system play an important role in forecasting the future trend of that system. In this paper, we present an approach to model the past crime behavior for future crime prediction. The study considered major crime event from Haryana state and used clustering approach to predict future crime trend. The analysis results obtained on ‘R’ tool for the past few years are found inconformity with that of real time trend, which envisage the success of our model proposed in this paper.

**Keywords** - crime location, criminal, crime prediction, clustering

## I. INTRODUCTION

Crime is an offence against the society. This unwanted anti-social behavior is punishable under law. It not only effect on the individual rather terrifically affect the people of the whole country. The law and prosecution agencies enshrined with the duty of mitigating the crime, adopts various measure to control the crimes in their realm. The crime events can be controlled effectively if the areas more prone to such events could be identified in advance. This quest for crime predictions make the issue more challenging and fetch the interest of research community on equal footing with that of security and law enforcement agencies.

The research community, especially those working in the area of Computer Science ponders over the issue from different angle and use mathematical approach to analysis of criminal behavior. One of the well-known examples of crime is attack of September 11, 2001 on world trade Centre in USA. Some other well-known crime examples are such as Nithari serial murder in Uttar Pradesh (2005-2006), serial terrorist attacked in Jaipur and Mumbai in 2008. All those are the crimes which are remains headlines in India for very long time because sympathy and sentiments are attached with the crime affected persons. The causality in these events could be minimized if an alert could be issues in advance. The computing research may bridge this gap between past crime events and advance alert. In this research paper we have tried to pave a way to bridge this gap. We have proposed a crime prediction model which could make a forecast for future occurrence of same event. This model is based on crime dataset of last few years obtained from National Crime Records Bureau (NCRB) for major cities in Haryana and predict for the possibility of happening of such event in the next year. Since the crime dataset have features resembling to Big Data. We have used platform of “R” tool to carry out analytical study. The following sub-sections of this paper describe the important terms and rationale behind this research.

### 1.1. Crime Behavior Prediction

Understanding the *modus operandi* crime phenomena is main activity of crime investigation agencies. The statistical evidences acquire critical importance when used to analyze the criminal behavior and inference could be drawn for future trend. The prediction may for similarity of crime types or may be regarding some geographical confinement of all or specific crime activities. All predictions serve as important input for security and law enforcement agencies to adopt the preventive measures and remain proactive in their mission.

### 1.2. Clustering Approaches for Crime Prediction

The technique of Clustering commonly used in data paradigm for grouping of objects or data based upon some common instinct. It does not have predefined features of their attributes rather work according to situation. Since present crime data have multi-attributes and semi structured format. Also it is huge dataset, hence Big Data approach serve well to cluster the crime dataset. Through such clustering meaningful pattern can be developed by tracing a relation between existing data. Here we have used the technique of nearest neighborhood to build a pattern. Applying clustering technique to finding similar crime trend on the basis of the multivariate crime attributes like that happened in Banks, Hospitals, Public offices and Shopping Centers etc.

The rest of this paper organized as follows: section 2, gives details of the work done by various technocrats and scientists in the field of crime prediction. Section 3 presents our proposed model of crime predictions. Section 4 gives the implementations and results. Section 5 concludes the paper and project for future scopes of this research.

## II. BACKGROUND

The field of crime predictions is a novice one and has much scopes to explore. However various researchers have

explored the are in depth. J. Xu Jennifer and Chen Hsinchun [1] proposed an efficient technique those shortest path algorithms to identify the paths association between the entities of the criminal networks. L. Buczak Anna and M. Gifford Christopher [2] using fuzzy association rule mining for crime pattern discovery. Vadivel A. and G. Shaila S. [3] discuss the Neuro-fuzzy model for crime event detection. Zeleznikow John et al. [4] proposed methodology developing decision support system for crime detection. J. Schreck Christopher et al. [5] examines the violent crime victims. S. Albanese Jay [6] provides a risk assessment would be taken to predict and control of organized crime. Bogomolov Andrey et al. [7] present an approach to predict crime in geographic space from sources of multiple data particularly on mobile phones.

Zhang Chao et al. [8] analysis of designing patrol allocation of opportunistic criminals. Poolsappasit Nayot and Ray Indrakshi [9] personalized the privacy on localitycreated services. Keppensa Jeroen and Schafer Burkhard [10] introduce novel knowledge driven methodology based on decision support system. Adderley Richard and Bond John [11] proposed a computer model to predict the probability of useful fingerprints, DNA, footwear marks of a volume crime. M. Arietta Sean et al. [12] provide method to identifying and validating automatically predictive relationship of visual appearance attributes. Maurushat Alana et al. [13] discuss criminal intelligence using big metadata. J. Xu Jennifer and Chen Hsinchun [14] provides framework for knowledge discovery for criminal network.

Kadar Cristina and Pletikosa Cvijikj Irena [15] developing city watch for crime prevention. Yu Chung-Hsien et al. [16] crime is classically unpredictable. They provide classification methods of data mining to predict future crime events. Calders Toon and Verwer Sicco [17] modify the naive Bayes classifier on a given sensitive attributes. Jaitman Laura and Machin Stephen [18] discuss about crime immigration. Farrell Graham et al. [19] legacies which include the crime drop and the adolescence role which is persistent offending. Murray Joseph et al. [20] gives childhood behavior to predict crime and violence in late adolescence. Wan Lee Sang et al. [21] proposed a learning framework to detect the patterns of human behavior on the sequential based actions. J. Agarwal, R. Nagpal and R. Sehgal in [22] have analyze crime and measured homicide crime captivating into account the subsequent year and that the trend is descending from 1990 to 2011. They have used the k-means clustering technique for extract useful information from the crime dataset using Rapid Miner tool. Priyanka Gera and Dr. Rajan Vohra in [23] have used a linear regression for forecast the happening of crimes in Delhi (India). They review a dataset of the last 7 years to predict amount of some crimes with murder, burglary, robbery and etc.

### III. PROPOSED MODEL

The main objective of this research paper is to determine the crucial features or characteristics that are contributing on higher level of criminal patterns or activities. The model is developed on various issues mentioned below:

- i) Identify those crime regions based on criminal characteristics.
- ii) Identify the variables which having high probability of crime occurrence.
- iii) Provide recommendations to security organization and crime fighting institution for better planning for crime prevention.

Table 1: Crime Variables for Structural Vectors

District	Murder	Robbery	Vehicle Theft	Burglary	Theft	Kidnapping
Ambala	14	13	267	394	171	3520
Bhiwani	1157	874	1463	6717	567	3123
Faridabad	982	763	1336	6121	525	1957
Fathead	991	711	1297	5129	461	900
Gurgaon	1062	638	9603	5011	488	733
Hisar	1005	734	8687	4445	496	714

The data collection are used in this paper are taken from the National Crime Records Bureau of India. It includes the police crime reports and attributes related to culture, transport, commerce, religion and government among others. Table 1 shows the structural vectors of the crime variable in which crimes of murder, robbery, vehicle theft, burglary, theft and kidnapping are recorded over X-axis. Another Y-axis attributes is districts of Haryana state in India.

Let  $C_1, C_2, C_3, \dots, C_p$  are crime events taken from data sources  $D_1, D_2, D_3, \dots, D_N$  and  $std(X)$  be standard deviation as calculated from dataset to identify the most deviating variable in the complete data set. The mapping of data sets would be optimized using big data tools

$$C_p: D_N \rightarrow O_B \dots \dots \dots (i)$$

where  $O_B$  is the optimized big data tool  
 $D_N$  is the crime data set and  
 $C_p$  is the objective of to predict crime

To persuade the objective a linear regression model is proposed by Wan Lee Sang et al [21] with intention to determine the relationship between dependent variable X and independent variable Y. First it will identify the structural vectors of criminal sites. The dependent variable is

the number of crime and subsequently independent variables is the attributes along with the studied area of crime. The model determines a coefficient with each attributes the predicted probability of high level of crime which is  $(p=1)$ . Finally the clustering technique is used to identifying their attribute wise crime trend. Figure 1 shows the flow chart of the crime analysis with flow of proposed methodology.

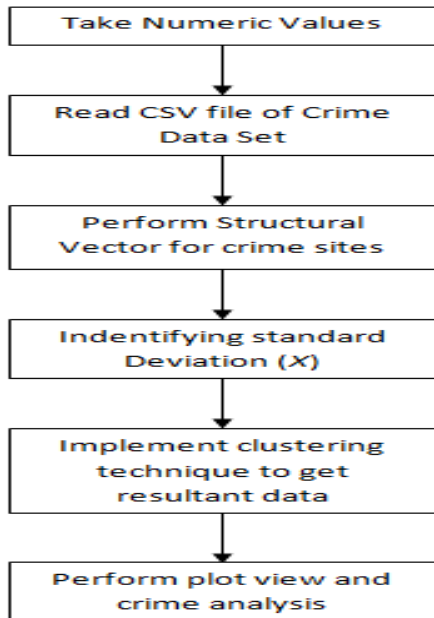


Figure 1: Crime Analysis Flowchart

The data analysis is carried out using district as a unit because police and law department are independent in their working at district level in India. Further, the state of Haryana is selected with the reason that it have diverse socio-demographic situation. To achieve the above objectives, clustering approach [22] is used in which similar pattern of crime is clustered into single identifying crime attributes for obtaining their crime pattern. The maximum number of crime happened in the attributes is likely to show the higher level of crime trends. Following procedure is adapted to calculation simulation results from proposed computational model:

- *Data as input.*
- *Nature of data as Numeric.*
- *A Structural Vector is associated with each Crime Site.*
- *Cluster of similar crime exists.*
- *Detection of crime attributes wise trend*

**IV. RESULTS**

Haryana The above proposed model is analyzed to determine the relationship between the attributes and their expected level of crime. The results of the above method are

fitted on the Big Data analytics ‘R’ tool in which data is analyzed and pattern is generated as output accordingly.

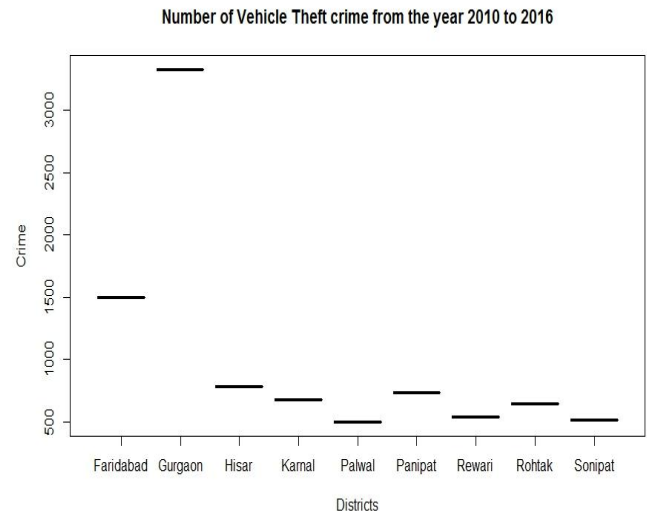


Figure 2: Vehicle Theft Crime District Wise

Figure 2 shows the number of vehicle theft crime from the year 2010 to 2016. Here X-axis defined as districts of the Haryana state with corresponding number of crimes. The data is taken as input and standard deviation is applied to get the crime site. It shows that the data is widely spread and need to analyses over crime sites. These districts having highest crime are showing by black line in figure 2.

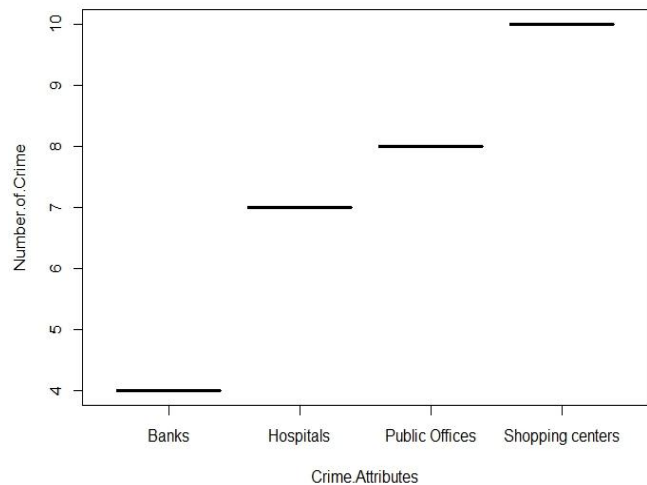


Figure 3: Attributes Wise Crime Trend

Figure 3 represents the attributes wise crime trend which includes shopping centers, public offices, hospitals and banks etc. These attributes are common places where theft occurred frequently. The clustering technique is applied to find similar crime trends. In which every districts having

crime attributes where crime is happened. So these crime attributes values are clustered of various districts of Haryana. Some attributes are having likelihood of more number of crimes and rest of having less. Similar crime patterns along with their crime attributes are shown in the figure 3.

## V. CONCLUSION AND FUTURE WORK

This paper concluded that the attributes like public offices and shopping centers have greater likelihood of a theft among other attributes like banks and hospitals. Such prediction will be helpful for the security departments to prevent theft in those critical areas. As a future work for the prediction of a crime model using more complex statistical methods and techniques to quantify the crime prediction.

## REFERENCES

- [1] J. Xu Jennifer and Chen Hsinchun, "Fighting organized crimes: using shortest-path algorithms to identify associations in criminal networks", *Decision Support Systems Elsevier*, pp. 473-487, 2004.
- [2] L. Buczak Anna and M. Gifford Christopher, "Fuzzy Association Rule Mining for Community Crime Pattern Discovery", *ACM*, 2010.
- [3] Vadivel A. and G. Shaila S. , "Event Pattern Analysis and Prediction at Sentence Level using Neuro-Fuzzy Model for Crime Event Detection", *Pattern Anal Applic Springer*, 2014.
- [4] Zeleznikow John et al, "A Methodology for Constructing Decision Support Systems for Crime Detection", *Springer*, pp. 823-829, 2005.
- [5] J. Schreck Christopher et al, "Examining What Makes Violent Crime Victims Unique: Extending Statistical Methods for Studying Specialization to the Analysis of Crime Victims", *J Quant Criminol Springer*, pp. 651-671, 2012.
- [6] S. Albanese Jay, "The prediction and control of organized crime: a risk assessment instrument for targeting law enforcement efforts", *Trends in Organized Crime*, pp. 4-29, 2001.
- [7] Bogomolov Andrey et al, "Once Upon a Crime: Towards Crime Prediction from Demographics and Mobile Data", *ACM*, pp. 427-434, 2014.
- [8] Zhang Chao et al, "Keeping Pace with Criminals: Designing Patrol Allocation against Adaptive Opportunistic Criminals", *International Foundation for Autonomous Agents and Multiagent Systems*, pp. 1351-1359, 2015.
- [9] Poolsappasit Nayot and Ray Indrakshi, "Towards Achieving Personalized Privacy for Location-Based Services", *Transactions on data privacy*, pp. 77-99, 2009.
- [10] Keppensa Jeroen and Schafer Burkhard, "Knowledge based crime scenario modelling", *Expert Systems with Applications Elsevier*, pp. 203-222, 2006.
- [11] Adderley Richard and Bond John, "Predicting crime scene attendance", *Policing: An International Journal of Police Strategies & Management*, Emerald Group Publishing Limited, pp. 292-305, 2008.
- [12] M. Arietta Sean et al, "City Forensics: Using Visual Elements to Predict Non-Visual City Attributes", *IEEE transactions on visualization and computer graphics*, pp. 2624 – 2633, 2014.
- [13] Maurushat Alana et al, "Using 'Big' Metadata for Criminal Intelligence: Understanding Limitations and Appropriate Safeguards", *ACM*, pp. 196-200, 2015.
- [14] J. Xu Jennifer and Chen Hsinchun, "CrimeNet Explorer: A Framework for Criminal Network Knowledge Discovery", *ACM Transactions on Information Systems*, pp. 201-226, 2005.
- [15] Kadar Cristina and Pletikosa Cvijikj Irena, "City Watch: The Personalized Crime Prevention Assistant", *ACM*, pp. 260-261, 2014.
- [16] Yu Chung-Hsien et al, "Crime Forecasting Using Data Mining Techniques", *Proceedings 11th IEEE International Conference on Data Mining Workshops*, pp. 779-786, 2011.
- [17] Calders Toon and Verwer Sicco, "Three naive Bayes approaches for discrimination-free classification", *Data Min Knowl Disc Springer*, pp. 277-292, 2010.
- [18] Jaitman Laura and Machin Stephen, "Crime and immigration: new evidence from England and Wales", *IZA Journal of Migration Springer*, pp. 2 – 19, 2013.
- [19] Farrell Graham et al, "Debuts and legacies: the crime drop and the role of adolescence-limited and persistent offending", *Crime Science a Springer Open Journal*, pp. 4-16, 2015.
- [20] Murray Joseph et al, "Childhood behaviour problems predict crime and violence in late adolescence: Brazilian and British birth cohort studies", *Soc Psychiatry PsychiatrEpidemiol Springer*, pp. 579-589, 2015.
- [21] Wan Lee Sang et al, "A Non supervised Learning Framework of Human Behavior Patterns Based on Sequential Actions", *IEEE transactions on knowledge and data engineering*, pp. 479-492, 2010.
- [22] J. Agarwal et al, "Crime analysis using k-means clustering", *International Journal of Computer Applications*, pp 1 - 4, 2013.
- [23] P. Gera, and R. Vohra, "Predicting Future Trends in City Crime Using Linear Regression", *IJCSMS*, pp 58 – 64, 2014.