

A Comprehensive Analysis of Crime Analysis Using Data Mining Techniques

S.Vijayarani¹, E.Suganya², C.Navya³

Assistant Professor¹, Ph.D Research Scholar², PG Student³

Department of Computer Science

Bharathiar University, Coimbatore, Tamilnadu, India 641046

¹vijimohan_2000@yahoo.com

²elasugan1992@gmail.com

³navyachandran24198@gmail.com

Abstract - *The crime is an act of doing something criminal. The crime rate and the cases are getting increased in recent past. Criminology is the study to understand the crime characteristic and the crime pattern. Due to the increasing crime rate, it is good to apply the data mining technique to identify the complexity of relationship between the criminals and crime pattern. The reports are collected from the National Crime Records Bureau. The records are written in the simple plain text format and it should be changed to a pattern or the format that can be used by the data mining tool. The data mining techniques are used to find the pattern and relationship in the given crime data set. This paper illustrates about the techniques and discussed about the recent related works that can be used to perform crime analysis.*

Keywords: Criminology; Personal Crimes; Property Crimes; Inchoate Crimes; Statutory Crimes; Cyber Crimes

I. INTRODUCTION

Crime is an unlawful action which is harmful to an individual and also to any society or the state. Criminology is a study which includes about crime and criminals. Analysis of crime is a function that involves an analysis for identifying trends and analyzing patterns in crime. Quantitative Analysis of data like regression analysis and linear programming and qualitative data analysis methods like content analysis, narrative analysis and grounded theory are used for the data analysis. Quantitative analysis of data is an approach to investigate data that is collected. The researcher then transforms the data that is observed or collected into numerical data. It often describes a situation by answering questions like 'what' and 'how many' questions about something. Qualitative Analysis of data is the range of procedures on the data that have collected to convert into some form of understandable explanation or interpretation about the people and situations that are being investigated. It is based on an interpretative philosophy. Crime analysis plays a major role in devising solution for the problems, and formulating crime forestalling strategies. Data mining techniques have higher influence in the fields such as law enforcement, narcotics, cybercrime, etc. The data mining techniques like clustering, classification, association rule mining can be used for the crime analysis and prediction.

The crime rate gets increased day by day. So it is important to know the crime factors and crime patterns. The data analysts of crime can help the crime officers to increase the speed of the process of resolving crimes with the newer techniques in data analysis. The data mining techniques like clustering and classification could be used get the pattern and factor of crime. The data mining techniques such as classification and clustering can be used to turn the raw data into some useful information. The crime data set is managed by National Crime Record Bureau (NCRB) to find the crime rate in a particular district or state. Any technique which helps to analyze the crime data easily and quickly is the efficient technique. The data in the crime data set should be changed into a format which can be understood by the mining tool and can use the mining techniques to understand the relationship between criminals.

II. RELATED WORKS

Year	Author Name	Title	Techniques	Inference
2018	Deepika K.K, Smitha Vinod [1]	Crime Analysis In India Using Data Mining Techniques	K-Means, Random forest, Neural networks.	The authors discussed about crime analysis of Indian states and union territories. Crime is analyzed using k-means clustering and classification is done using random forest algorithm and neural networks. Visualization is achieved using Google marker.
2018	AyisheshimAlma w, KalyaniKadam [2]	Survey Paper On Crime Prediction Using Ensemble Approach.	Naïve Bayes, J48, Artificial network.	The authors described about investigation of data mining techniques and group classification techniques for discovery and prediction of upcoming crime.
2018	Dr. M. Sreedevi, A.HarshaVardhan Reddy, Ch. VenkataSaiKrishna Reddy [3]	Review On Crime Analysis And Prediction Using Data Mining Techniques	Data Mining, Machine learning, Crime analysis, Crime prediction.	The authors have discussed about the filtering of data set using the genetic algorithm and classification has done using decision tree using GINI index.
2017	GouravGovindaswamy, Vinod Kumar Kathineni, Santhosh Kumar P [4]	A Survey On Crime Data A Analysis Using Data Mining Techniques	Data Mining, Crime data analysis, Knowledge discovery process (KDD), Neural Networks.	The authors explained the data mining techniques like classification like neural networks and decision making, clustering, prediction and association rule and its role on crime application.
2017	PrajaktaYerupude, VaishnaviGudur [5]	Predictive Modeling Of Crime Dataset Using Data Mining	Supervised learning, Unsupervised learning, Decision tree, Naive Bayes, Regression, Data mining, Machine learning.	The authors explained the algorithms like decision tree, Naïve Bayes are applied on the data set to predict features that affect the high crime rate. Supervised and unsupervised learning are used to get the output.
2017	SurpreetKaur, Dr. Williamjeet Sing [6]	Systematic Review Of Crime Data Mining	Crime Data Mining, Crime Data Analysis	The authors discussed about the clustering, classification, association rule mining, frequent pattern mining, outlier analysis and regression to find the matching and frequent crimes.
2016	ArnabSamanta, AmolJoglekar [7]	Crime Classification and Criminal Psychology Analysis using Data Mining	Crime Analysis, Data Mining, Classification Rules, Clustering.	The authors proposed a model for classifying crimes based on the level of seriousness and visualization. It also provide a functionality for analyzing the psychology of murderers
2015	TahaniAlmanie, RshaMirza, Elizabeth Lor [8]	Crime Prediction Based on Crime Types and Using Spatial and Temporal Criminal Hotspots	Data Mining, Crime Prediction, Crime Classification, Crime Frequent Pattern, Denver & Los Angeles Criminal hotspot	The authors focused on finding temporal and spatial crime hotspot using two real-world crime datasets and provides comparison between the datasets. The result of this solution can be used for awareness to the people and it help the agencies to predict the future crimes.

2015	TusharSonawanev , ShirinShaikh, ShaistaShaikh, Rahul Shinde, AsifSayyad [9]	Crime Pattern Analysis, Visualization And Prediction Using Data Mining	K-Means, Cluster, Correlation.	The authors described about the crimes against women. The dataset is classified based on the predefined conditions. It used K-means algorithm for classification and also correlate crime and predict the crime.
2013	Anshu Sharma, Raman Kumar [10]	Analysis And Design of An Algorithm Using Data Mining Techniques For Matching And Predicting Crime	Data mining, Criminology, Clustering, Classification.	The authors explained about classification techniques like decision tree and clustering techniques like k-means to get pattern of crime data and also K-means weighted approach. The performance measure of the proposed k-means algorithm comes out to be 0.989, 0.987 and 0.987.
2012	Shashishekhar, Pradeep Mohan, Dev Oliver, XunZohu [11]	Crime Pattern Analysis: A Spatial Frequent Pattern Mining Approach	Spatial Frequent Pattern Mining, Regional Frequent Crime Pattern, Crime Outbreak Detection.	The authors presented a case study to discover interesting, useful and non-trivial crime outbreaks in a dataset from Lincoln, NE. A review of emerging trends in crime outbreak detection is also presented.
2006	ShyamVaranNath [12]	Crime Pattern Detection Using Data Mining	Crime Patterns, K-means, Law Enforcement, Semi-supervised learning,	The authors discussed about the use of clustering algorithm for detecting the crime pattern and used semi-supervised learning for knowledge discovery.

III. TYPES OF CRIMES

Crimes can be divided into five major categories: personal crimes, property crimes, inchoate crimes, statutory crimes and cybercrimes.

A. Personal Crimes

It is defined as offenses against a person. The personal crime is a misuse or harm (physically or mentally) to some another person. It includes assault, kidnapping, homicide, rape, fault imprisonment, etc [7]

- **Assault** - Assault is an act of physical harm or unnecessary physical contact with a person in some specific illegal definitions or a trial to commit such crimes.
- **Kidnaping** - In criminal law, kidnapping is the unlawful transportation of a person or controlling a person against their wish. It can be a composite crime.

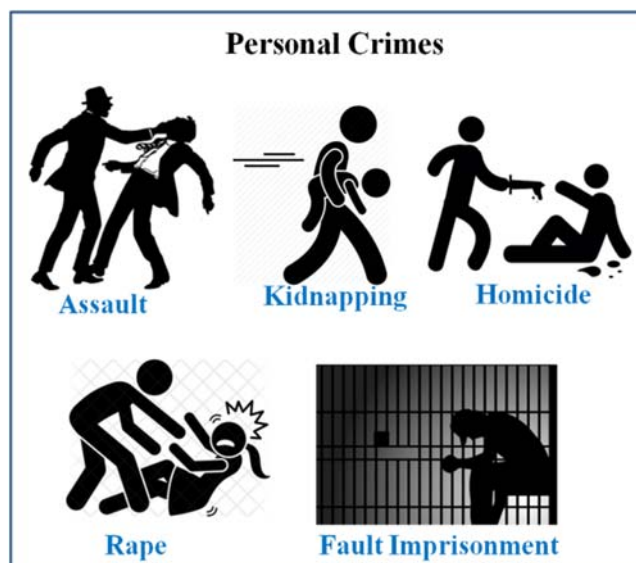


Figure1. Personal Crimes

- **Homicide** - Homicide is the act of one human killing another. A homicide needs a desired act by a person that ends in death. A homicide may result from accidental or inattentive acts if there is no intention to cause any harm.
- **Rape** - Rape is an act of intercourse sexually with a person without their permission by force or intimidation of force. In many jurisdictions rape has categorized under the sexual assault, which also encloses the acts that fall in short of intercourse.
- **Fault Imprisonment** - False imprisonment is prohibiting a person in a surrounded area without any justification or consent. False imprisonment is a common law offence and a wrongdoing.

B. Property Crimes

It is defined as offenses against the property. Property crime does not involve any physical harm to another person. Instead of that, it involves a conflict with a person's right to use their own property. Property crimes include forgery, robbery, burglary, arson, theft, embezzlement, false pretenses.



Figure2. Property Crimes

- **Theft** - Theft is getting one's property without his or her permission or with the desire to rob from the right owner of that property. The theft is also referred as filching, thieving and stealing.
- **Robbery** - Robbery is a type of property crime that attempts to take any property or service that has value by force or threatening or scaring the victim. In common law, robbery is defined as acquiring the belongings of another person, with the intention to permanently rob the property by force or fear.
- **Burglary** - Burglary, also called as entering or breaking or sometimes housebreaking. It is an illegal entry into other's building or location with the objective of committing an offence.
- **Arson** - Arson is the crime of intentionally and caustically setting fire to charring property. A common motive for arson is to commit insurance fraud.
- **Forgery** - Forgery is an executive crime that commonly used to refer to the making or alteration of material of a legal instrument with the specific intention to deceive someone.
- **Embezzlement** - Embezzlement is the act of prohibiting any for the purpose of conversation of those assets, by any one or more people that the assets were handed over, either to keep it safe or to use for some purposes with some conditions.
- **False Pretenses** - False pretense is referred as misrepresenting a property with an illegal evidence in order to have someone's property.

C. Inchoate Crimes

Crimes that were started, but not finished are called as inchoate crimes. Inchoate crimes contains that a person to take a considerable step to finish the started crime, as opposed to just intend to carry out a crime. It includes conspiracy, solicitation, attempt, incitement, etc.



Figure3. Inchoate Crimes

- **Attempt** - An attempt refers to a crime that occurs if a criminal has intention to commit a crime and takes the entire necessary steps for completing the crime. It is a plan to execute a crime but in the last the planned crime does not occur.
- **Solicitation** - Solicitation is the act of offering, or purchasing goods or services illegally. The legal status is specified to the time or place where the crime has occurred.
- **Conspiracy** - A conspiracy is an arrangement between groups of persons to commit a crime in the future at some time.
- **Incitement** - Incitement is the assistance of another person to execute a crime. In common law, some types of incitement acts may be illegal.

D. Statutory Crimes

A violation of a specific rule or federal regulation and can be involved in either property or personal offenses. Three significant types of statutory crimes are alcohol related crimes, drug crimes, traffic offenses, and financial/white collar crimes. Statutory crimes include drunk and drive, intoxication, minor in-possession of alcohol and refusing sobriety testing.

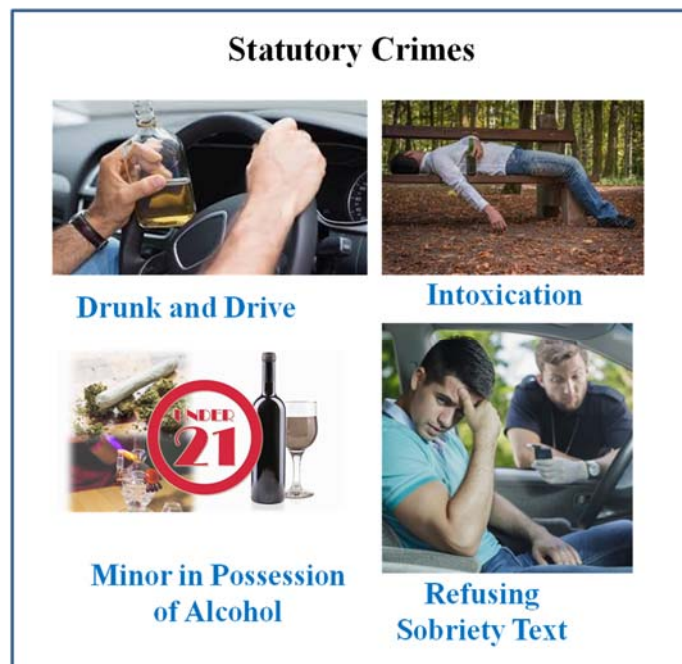


Figure4. Statutory Crimes

- **Drunk and Drive** - Driving under the influence (DUI) of alcohol or drugs is a crime. It is an offence of driving or handling a vehicle after the consumption of alcohol or drugs, to a level above the capability of driver for operating their vehicle.
- **Intoxication** - Intoxication is a subsided ability to act because of any effect of chemical substances consumption. In criminal law, it is an offense if a person claim diminished responsibility of substance intoxication.
- **Minor in Possession of Alcohol** - It is a violation that arises when a person below the lawful drinking age of 21 is found as using alcohol.
- **Refusing Sobriety test** - Sobriety tests are free-willed and fulfillment of the tests is not necessary under the law. But if a driver refuses a field sobriety test, he or she will be asked to undertake a chemical test to determine his or her blood alcohol level.

E. Cybercrimes

Cybercrime is also called as computer crime. It defined as using a computer systems as an equipment to carry out illegal acts such as stealing identities, trafficking and child pornography, committing fraud intellectual property, and violating privacy. Cybercrime includes hacking, child pornography, virus dissemination, computer vandalism, cyber terrorism and software piracy and denial of service.



Figure5. Cyber Crimes

- **Hacking** - Hacking is an attack to utilize a computer or a private network inside any computer system. It is the unauthorized access to take control over computer network systems for some illegal mission.
- **Child Pornography** - Child pornography is known as any representation of a minor or a person who looks like a minor that is hooked in sexual conduct or sexually related conduct. This can be images, videos or any digital data that are generated using a computer system.
- **Virus Dissemination** - It is a process of vengeful software that ruins the system of any other person. They disturb the operation of computer and affect the data by changing or deleting it.
- **Computer Vandalism** - Computer vandalism is a program that performs awful operations such as deriving any user passwords or some other data or wiping out the hard drive.
- **Cyber Terrorism** - Cyber terrorism is considered as the politically provoked use of information technologies and computer networks to generate caustic disruption or to spread fear in society.
- **Software Piracy** - Software piracy is the action of pirating or stealing any software which is legally secured. This stealing includes selling, distributing, copying or modifying the software.
- **Denial of Service** - It is a type of cyber-attack in that the criminal seeks to set a system or network unavailable to its original users temporarily or indefinitely by disrupting services from a host that is used to connect to the internet.

IV. CRIME ANALYSIS METHODOLOGY

Crime analysis and prediction methodology has four technologies. Four techniques can be used in the field of crime analysis. The crime analysis methodology is as follows:

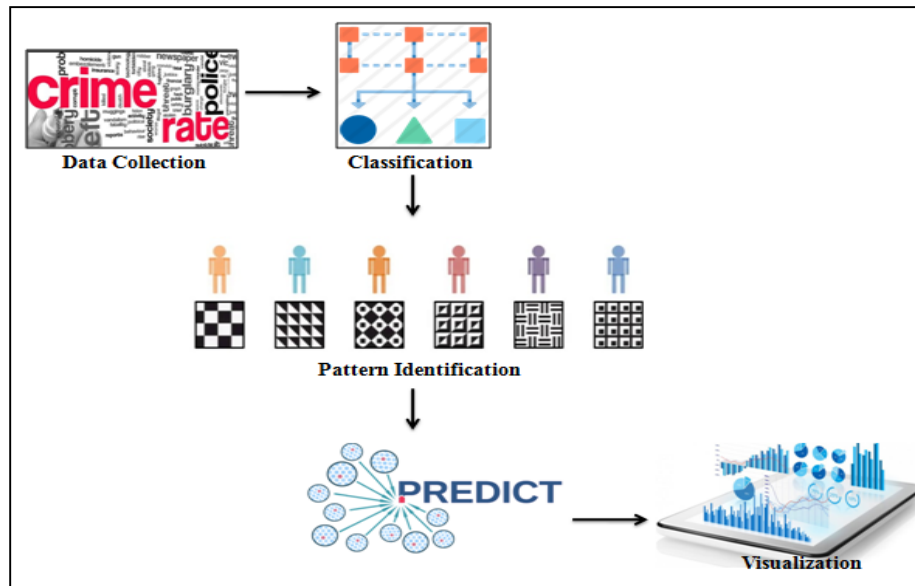


Figure 2: Systematic Methodology of Crime Analysis

A. Data Collection

Data sets are collected from various sources. The reports are gathered from the department of police records [10]. Data collection is done at the end of year at police reports which made available by NCRB [9]. Next the records are collected from resources like new sites, blogs and social media [3]. The records also collected from National Crime Record Bureau (NCRB) and police department [5]. Data collection is also done with the police department reports, previous investigation files, intelligence reports, open source intelligence findings and police arrest records. The data collected will be in many formats such as text, graph images which are called as unstructured data and relational data and also as semi-structured data. The collected data can be stored as a database for further processing. Table shown the some of the link used for data collection:

Table 1. Data Collection

S.No	Crime Dataset Repository	URL
1.	National Crime Record Bureau(NCRB)	https://ncrb.gov.in/
2.	Data World	https://data.world/datasets
3.	UCI Repository	http://archive.ics.uci.edu/ml/datasets
4.	Kaggle	https://www.kaggle.com
5.	Data.gov in	https://data.gov.in
6.	Knoema	https://knoema.com
7.	India open data census	https://in-city.census.okfn.org/dataset
8.	Springboard blog	https://springboard.com
9.	Socrata	https://www.tylertech.com/products/socrata

B. Data Classification

Classification technique is used to identify the pattern of the crime data. Naïve Bayes algorithm is used for the supervised learning method. The concept of Named Entity Recognition used to classify all the elements in the text into some predefined categories like names, locations and so on [3]. The random forest [5] and K-Means algorithm [9] is also used for classification and clustering. The main application of classification is fraud detection and credit card services. The decision tree and neural network based classification algorithms [4] are used for classifying the criminal data. Arnab Samanta et al. has proposed a new model to analyze the crimes. The authors focused only psychology of murders. There are two stages of their proposed model: first classified the crimes then the analyzing the psychology of murders. The classification of crimes is different kinds: Murder, Rape, Theft, Traffic Violation, Kidnapping, Cyber-crime, Assault, Trespassing and Vandalism. The authors proposed a new model which is based on the level of seriousness of the crime. Hence the authors classified the crimes into three classes that are Felonies, Misdemeanors, and Violation using rules which is named as PRISM.

C. Pattern Identification

Identification of pattern is used to find the trends and patterns in crime using various data mining technologies like classification and clustering. Clustering techniques are applied on the existing and known crimes. Apriori algorithm determines the crime pattern and association rule [3] is used to identify the particular place in which the crime rate has increased. Pattern identification avoids the crime occurrence in a particular place by providing security, CCTV, fixing alarms etc. Artificial neural network is a biological system that detects the crime patterns [4]. The usage of K-means clustering [1] techniques forms the cluster based on the similarity of crime pattern.

Shyam Varan Nath[12] has implemented machine learning framework for geo-spatial plotting of crime. It helps for improving the capacity of the private investigators and other law enforcement officers. It can be adapted for counter terrorism for homeland security. The authors collected the data from a Sherriff’s office, under non-disclosure agreements from the crime reporting system. The operational data was converted into denormalized data using the extraction and transformation [12]. The author identified seven patterns from the dataset using clustering data mining technique. In this figure 6 [12] shows patterns which are based on the crimes. The authors identified seven patterns and 309 crimes using data mining technique.

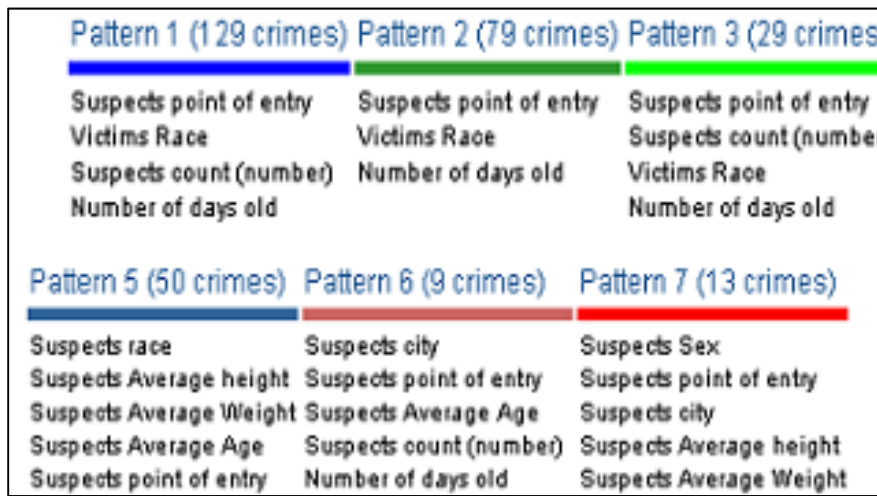


Figure 6: Pattern Identification [9]

D. Prediction

Prediction is used to state the probability of an event in future. Decision tree concept [3] is used for the prediction of future crime by using the larger data sets. It is a supervised machine learning techniques. Weighted K-means algorithm [10] is also used for prediction. Predictive analysis is a statistical methodology used to evolve models that could be used to predict future events. Prediction can be done with the techniques like decision trees, naïve bayes and linear regression [5]. The artificial neural network [4] is also used to make prediction. Regression techniques can be adapted for prediction and neural networks can create regression.

Tahani Almani et al. focused finding spatial and temporal criminal hotspot. They analyzed two different real time datasets: Denver and Los Angeles using Apriori algorithm. Denver dataset contains 19 attributes and 333068 instances. It has categories such as robbery, public-disorder, and sexual assault. Los Angeles dataset contains 14 attributes and 243750 instances. It has details about Theft-Plain, Theft-Person, and Theft-From-Motor-Vehicle. The authors implemented Bayesian classifier to forecast an expected crime type. They formed the time interval from T1 to T6 and predicted the crime type based on the crime features.

Frequent pattern	Min-sup	Frequent pattern	Min-sup
'Capitol-hill', 'Monday', 'T5'	0.001	'Five-points', 'Thursday', 'T4'	0.001
'Capitol-hill', 'Thursday', 'T6'	0.001	'Five-points', 'Thursday', 'T5'	0.002
'Capitol-hill', 'Friday', 'T5'	0.001	'Five-points', 'Thursday', 'T6'	0.002
'Capitol-hill', 'Friday', 'T6'	0.002	'Five-points', 'Wednesday', 'T3'	0.001
'Capitol-hill', 'Saturday', 'T6'	0.002	'Five-points', 'Wednesday', 'T4'	0.002
'Capitol-hill', 'Sunday', 'T6'	0.001	'Five-points', 'Wednesday', 'T5'	0.002
'CBD', 'Monday', 'T4'	0.001	'Five-points', 'Wednesday', 'T6'	0.002
'CBD', 'Monday', 'T5'	0.001	'Five-points', 'Saturday', 'T1'	0.001
'CBD', 'Tuesday', 'T3'	0.001	'Five-points', 'Saturday', 'T5'	0.002
'CBD', 'Tuesday', 'T4'	0.001	'Five-points', 'Saturday', 'T6'	0.002
'CBD', 'Wednesday', 'T3'	0.001	'Five-points', 'Sunday', 'T1'	0.001
'CBD', 'Wednesday', 'T4'	0.002	'Five-points', 'Sunday', 'T5'	0.001
'CBD', 'Wednesday', 'T5'	0.001	'Five-points', 'Sunday', 'T6'	0.002
'CBD', 'Wednesday', 'T6'	0.001	'Montebello', 'Monday', 'T6'	0.001
'CBD', 'Thursday', 'T3'	0.001	'Montebello', 'Wednesday', 'T6'	0.001
'CBD', 'Thursday', 'T4'	0.001	'Montebello', 'Thursday', 'T6'	0.001
'CBD', 'Thursday', 'T5'	0.002	'Montebello', 'Friday', 'T5'	0.001
'CBD', 'Friday', 'T4'	0.001	'Montebello', 'Friday', 'T6'	0.001
'CBD', 'Friday', 'T5'	0.001	'Montebello', 'Saturday', 'T6'	0.002
'CBD', 'Friday', 'T6'	0.001	'Montebello', 'Sunday', 'T6'	0.002
'CBD', 'Saturday', 'T5'	0.002	'Stapleton', 'Wednesday', 'T5'	0.001
'CBD', 'Saturday', 'T6'	0.002	'Stapleton', 'Friday', 'T5'	0.002
'Five-points', 'Monday', 'T5'	0.002	'Union-station', 'Friday', 'T6'	0.001
			0.002

Figure 7. Predicted values of Given Dataset

In the figure 7 shows the predicted values of Denver and Los Angeles dataset using Apriori algorithm. This figure is taken from Anshu Sharma et al. [10].

E. Visualization

Visualization is the process of representing the extracted knowledge or result. The visualization can be in the form of graphs, bar-concepts, pie-charts, line graph and so on. The authors mostly used area chart and scatter plot for graph representation of predicted crimes based on the location. Figure 8 shows the chart representation of crimes and its types [11]. This figure is taken from Shashi Shekhar et al. [11].

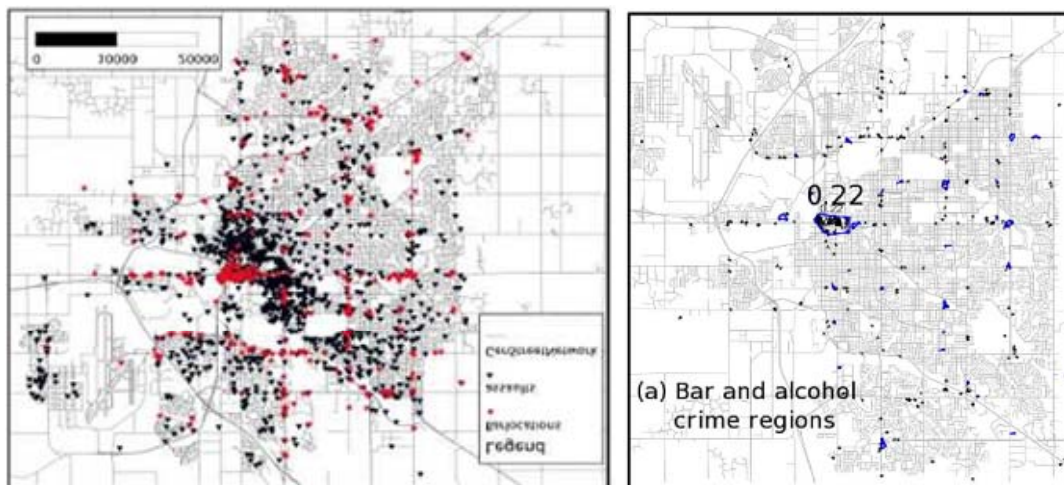


Figure 8. Visualization

V. RESEARCH ISSUES AND CHALLENGES

- Crime analysis using data mining has many issues and challenges in collection, preprocessing, storing and visualization has its own issues.
- Handling of relational and complex types of data is an issue.
- The main challenge is to handle the data that have different structure or format and selection of appropriate technologies and techniques.
- The collection of data set from various heterogeneous and autonomous sources like social media, criminal records, police reports are challenging task.
- The collected data can be of different formats like text, graph, images, unstructured and semi-structured data. Transforming different types of data is another challenging process.
- The crime data can be of larger volume of data storage management is another tedious process

VI. CONCLUSION

Crime analysis and pattern prediction is important in today's environment. The combination of facts such as extensive growth of terrorism and the lack of truly secure system makes it an important field of research. Data mining is applied in the context of law enforcement and crime related problem. Wide range of people, society regions and world is affected with crime. Crime prediction helps the people stay away from the districts and improve the living style. In addition, having this kind of knowledge would help the people in travelling choice places. The result of this analysis helps users in understanding the range of available crime data mining techniques and technologies. This paper discussed about the basic concepts of crime analysis, its process, existing methods and research challenges. This review will be useful for researchers who all are doing research under crime analysis and also to the department of police and crime branch officers and crime investigators.

REFERENCES

- [1] Deepika K.K, SmithaVinod, "Crime Analysis In India Using Data Mining Techniques", International Journal of Engineering & Technology, 7 (2.6) (2018).
- [2] AyisheshimAlmaw, KalyaniKadam, "Survey Paper On Crime Prediction Using Ensemble Approach", International Journal of Pure and Applied Mathematics, Volume 118 No-8.
- [3] Dr. M. Sreedevi, A. HarshaVardhan Reddy, Ch. VenakataSai Krishna Reddy, "Review On Crime Analysis And Prediction Using Data Mining Techniques", International Journal of Innovative Research in Science, Engineering and Technology, Vol 7, Issue 4, April 2018.
- [4] GouravGovindaswamy, Vinod Kumar Kethineni, Santhosh Kumar P, "A Survey on Crime Data Analysis Using Data Mining Techniques", Int. Journal of Engineering Research and Application, Vol. 7, Issue 8, August 2017.
- [5] PankajakshaYerupude and VaishnaviGudur, "Predictive Modeling Of Crime Dataset Using Data Mining", International Journal of Data Mining & Knowledge Management Process (IJDMP), Vol. 7, No. 4, July 2017.
- [6] SarpreetKaur, Dr. Williamjeet Sing, "Systematic Review of Crime Data Mining", International Journal of Advanced Research in Computer Science, Volume 8, No. 5, June 2017.
- [7] ArnabSamanta, AmolJoglekar, "Crime Classification and Criminal Psychology Analysis using Data Mining", International Journal of Mechanical Engineering and Information Technology, Vol-04, Issue 10, October 2016.
- [8] TahaniAlmanie, RshaMirza and Elizabeth Lor, "Crime Prediction based on Crime Types and using Spatial and Temporal Criminal Hotspots", International Journal of Data Mining & Knowledge Management Process (IJDMP) Vol.5, No.4, July 2015.
- [9] TusharSonawanev, ShirinShaikh, ShaistaShaikh, RahulShinde, AsifSayyad, "Crime Pattern Analysis, Visualization And Prediction Using Data Mining", IJARIE, Vol-1, Issue-4, 2015.
- [10] Anshu Sharma, Raman Kumar, "Analysis And Design Of An Algorithm Using Data Mining Techniques For Matching And Predicting Crime", International Journal of Computer Science and Technology, Vol-4, Issue 2, June 2013.
- [11] ShashiShekhar, Pradeep Mohan, Dev Oliver, XunZohu, "Crime Pattern Analysis: A Spatial Frequent Pattern Mining Approach", May 2012.
- [12] ShyamVaranNath, "Crime Pattern Detection Using Data Mining," 2006, IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology Workshops, Hong Kong, 2006, pp. 41-44.doi: 10.1109/WI-IATW.2006.55