Crime Analysis and Prediction Using Enhanced Arima Model

S. Vijayarani[1], E. Suganya[2], C.Navya[3]

[1]Assistant Professor, Department of Computer Science, Bharathiar University, Coimbatore ,INDIA.

[2] Ph.D Research Scholar , Department of Computer Science, Bharathiar University, Coimbatore, INDIA.

[3] M.Sc Student , Department of Computer Science, Bharathiar University, Coimbatore, INDIA.

A B S T R A C T

Crimes are known as unforeseeable and big social nuisances that can cost our society in many ways. There are many types of crimes arises every day and it leads to many new trends in crimes. Criminology is a scientific study, which combines the crime characteristics, crime pattern, and criminal behaviors. Most representative researches include predicting crimes, crime pattern discovery, crime hot-spot methods, relationship between crime and meteorological and temporal factors etc. The fast technological development as well as the newly invented technologies has created many new possibilities in crime analysis. The complexity in handling the huge amount of data produced each day has made it as an important field to apply data mining techniques. Crime analysis place one of the important roles in crime solving. In this paper, we make use of time series analysis on Chicago crime data. The intention of this paper is to understand the main use of time series algorithms to find the top crimes of Chicago on monthly, weekly, and daily basis. One can create more knowledge from the past data, help in reducing the crime count, and ensure security.

Keywords: Crime Prediction, Illinois Compiled Statutes (ILCS), Time Series Analysis, ARIMA

1. Introduction

The crime rate is increasing considerably day by day. It is a disastrous act for humanity and an obstacle in the way of development. A crime defines an action that constitutes an offense that may be prosecuted by the police and is punishable by the law. Crimes mainly divided into four categories such as property crimes, personal crimes, statutory crimes and inchoate crimes [12, 4]. Apart from this many other crimes such as cybercrimes, crimes against animals and nature are also there. People of all age groups commit crime. Earlier only males were reported committing crimes whereas now reports prove that both women and men commit crimes [4]. Crimes like sexual harassments, robberies; looting, abduction, rape and killing are some of the major crimes that are happening everywhere. As these crimes are lifting high, the need to control such crimes has been creating a huge pressure on investigation department. With the increase in the use of computer systems to trace crimes and track criminals, data analyst has started helping the detectives as well as the law officers to speed up the investigation and other processes.

Data mining is defined as the process of uncovering the hidden pattern or extracting the hidden data. Data mining has two categories, which includes classification with prediction and description mining. Data mining contains many techniques which includes classification, clustering [7] association rule mining, frequent pattern analysis, outlier analysis, regression and predictions and each one of them having its own application and importance. The

* *Corresponding author.*
E-mail address: elasugan1992@gmail.com

extraction of data from the large volume of data has occurred for centuries. Collection of data, storage of the data and manipulations of data have been increased because of the proliferation and the increased power of increased technology.

Crime data mining involves exploiting data about crimes to enable law enforcement to better apprehend criminals and prevent crimes. Analyzing this data not only helps in recognizing a feature responsible for high crime rate but also helps in taking necessary actions for preventing the crimes. Crime is a behavior disorder that is integrated result of social, economic and environmental factors [19]. The need to analyze the crimes is increasing day by day. It helps the law enforcers to grab the details about crime pattern, trend, series and timely manner and also helps understand the need of the changing society. Analyze the crime help to utilize the prosperity of information existing in any public domain, system of justice and law enforcement agencies. Analyzing crimes also leads to increase the use of law enforcement resources and understanding criminal behaviors and to find proactive way in detection and prevention of crime [17]. Some of the most important researches that take place in crime data mining are hot-spot method, relationship between crime and meteorological and temporal factors, prediction of crime, crime pattern on vacation, factors that influence crime rate, time series analysis etc.

Crimes cannot be predicted since it is not systematic or random [16]. But different classification algorithm in problem resolving and time series analysis is nominated based upon the suitable requirements in crime data prediction. Each technique provides a diverse accuracy and estimated outcome [3]. Analyzing the crime data not only helps in predicting crime but also responsible for recognizing a factor that are reasons for high crime rate. At large crime forecasting consider some parameters as response, forces used, shooting and injuries, and the factors that are related to such parameters. In this context, forecasting using time series has become a highly significant approach to aid in such situations [21]. The standard time series approach to intervention analysis is mainly based on autoregressive integrated moving average (ARIMA) models. Structural time series may provide an effective approach to intervention modeling [20].

Each day each minute a crime is occurring. It produces a large volume of data. So handling such data is a challenging process also. Handling criminal data is generated by the law enforcement organizations is a major challenge for them to analyze this data to implement decision for avoiding crimes in future. Even if a small volume of data is collected, the occurrences of crime patterns remain complex [1]. The investigation may take a lot of time because of such complex issues. Another challenge is to find varying trends and solving crimes has been the prerogative of the criminal justice and law enforcement specialists. Another major issue with the crime data is storage and the format of the data. Increasing of crime information should be stored in a well-organized manner to analyze. Crime data taken for some other uses other than analysis may vary in formats and may contain some missing values and it is one of the issues in crime data mining. However, the preprocessing techniques applied on crime data for better accuracy. The remaining part of the paper is organized as follows: Section II explains few important research approaches in crime analysis using data mining techniques and the approaches, which use time series analysis for forecasting crimes. Section III represents the proposed methodology and theory. Section IV delivers the information regarding the top five crimes analyzed in this paper and section V contains the results of our model. Finally, section VI concludes the paper.

2. Related Works

We have referred many papers that proposes many types of methods as time series analysis, crime prediction, crime hot-spot identification, criminal behavior analysis, finding crime trends etc. some of the papers are discussed here.

Sudan Jha et al. [1] used a strategy for a time series model and machine testing systems for crime estimation. They have used existing machine learning techniques for prediction of crime count and proposed two approaches such as modified autoregressive integrated moving average model and a modified artificial neural network. The paper concludes that using a time series model yields better forecasting results than the predicted values from existing technologies.

Tahani Almanieet al. [15] are focused to find criminal hotspots. They made use of Apriori algorithm for the production of interesting frequent pattern to observe the criminal hotspot and they used decision tree and naïve Bayesian classifier for predicting potential crime categories. They applied the algorithms for Denver crime data and Los Angles crime data. They combined the findings of Denver crimes datasets with demographics information.

Arnab Samantaet al. [12] proposed a model, which is used to classify crimes, based on their level of seriousness and will provide visualization. The model also provides functionality for analyzing psychology of murders using clustering technique. The classification is used to classify the crimes based on the level of seriousness and a sequential covering algorithm called PRISM is used to learn rules for classification from the training set. Clustering is performed for analyzing the psychology of murders and a variation of K-Medoids algorithm called as CLARANS algorithm is used for clustering.

Md. Abdul Awalet al. [8] uses linear regression algorithm to predict future crime trends in Bangladesh. The algorithm helped in forecasting the crimes such as kidnapping, women and child repression, robbery, burglary, dacoit theft and murder. The result has been categorized into two parts such as metropolitan region result and divisional region result for better and easier understanding. The paper is concluded from the result as most of the crimes are increasing with the growth of population.

Quanbao Jianget al. [21] comes with stochastic crime rate forecast method namely based on the historical age-specific crime rates. They have employed SVO (Singular Value Decomposition) in matrix theory to lower the rank of the crime rate matrix. This paper provides a method for forecasting age-specific crime rate by age using singular value decomposition and time series modeling. They used the age-specific crime rate as a foundation, the crude crime rate can be determined through population forecast model and future crude crime rate by decomposing the difference between two crude crime rate.

Grzegorz Borowiket al. [5] performed time series analysis of crime on the Republic of Poland involving about 1.2 million crime events. The purpose of this paper is to show the usefulness of analytics algorithm in predicting crimes. The study presents the results of six types of crime hooliganism, police interventions, burglary, road-traffic offenses, theft and detention. The experimental results has been obtained by applying spectral analysis and prophet forecasting to the time series of actual crime and public offence counts in Poland confirm the utility of crime time series analysis.

Md Abu Salehet al. [2] used machine learning K-means algorithm for prediction and to analyze the crime in Chicago town and the dataset is accessible at Kaggle website. The next intention of this project is to check the feasibility of the k-means algorithm to determine and solve the crime problems.

Attributes focused for analysis are type, time, location of the crime and arrest made. The result has been represented through bar graph, histogram and pie-chart.

Sanjeev Sridharanet al. [26] analyzes the legislation impact on the rate of reported crimes using variety of time series approaches. Structural time series models are designed instead of box-jenkins ARIMA model. The structural time series framework provides a broader framework and allows the parameter to change over the time period. The affability of structural time series is embossed by representing a variety of analysis that give additional information about the deterrent impact of the new legislation.

Suncica Vuji´cet al. [20] used mainly a structural time series analysis to dig into the bang of parole abolition and sentence reformation in the city of Virginia on the rate of reported crimes. This study for the city of Virginia also gives an illustration about the effectiveness of the time series analysis when used in crime studies. The results have been notified that the legislation of January 1, 1995 have been affected burglary (-3.99) and murder (-0.09).

Dr: Zakaria et al. [17] proposed a model for crime data and criminal and analyses using simple K-means algorithm for clustering and Apriori algorithm association rules. The goal of this project is to help the government of Libya to take strategic decision about prevention in the increasing crime rates. They have used WEKA software and Excel software to preprocess and analyze the collected crime data as well as criminal data. K-means clustering algorithm gives the overall statistical knowledge the criminal age and crime type.

Deepika K.Ket al. [4]includes steps such as data pre-processing, clustering, classification and visualization. Crime is identified using k-means clustering and the clusters are formed based in the similarity of the crime attributes. The Random forest algorithm and neural networks are applied on the data for classification. Visualization is achieved using the Google marker clustering and the accuracy is verified using WEKA tool. The proposed model is very useful for both the investigating agencies and the police officials in taking necessary steps to reduce crime. The model can be applied to any countries dataset.

3. Methodology

In this paper, time series model of ARIMA is used to make forecasting of top five crimes namely theft, battery, criminal damage, assault and deceptive practice for the crime data of Chicago.The method has four stages: identification, estimation, diagnostics checking, and forecasting [1]. The ARIMA method is useful for analyzing longitudinal data with a correlation among neighboring observations and analyzing multivariate time series.

ARIMA (p, d, q) model is a linear model that fit for dealing with stochastic series. It is originated from the autoregressive model AR (p), the moving average model MA (q) and the combination of AR (p) and MA (q), the ARMA (p, q) model [24].

The ARIMA methodology, developed by Box-Jenkins, focus on the idea that series can be made stationary by operations such as differencing.In the case that the time series is difference stationary, it might be transformed to stationary by taking first or second differences possibly in combination with seasonal differencing. Formally we define first and second differences by

$$\Delta y_t = y_t - y_{t-1} \qquad\qquad (1)$$
$$\Delta^2 y_t = \Delta y_t - \Delta y_{t-1} = y_t - 2y_{t-1} + y_{t-2} \qquad (2)$$

respectively, and seasonal differences by

$$\Delta_s = y_t - y_{t-s} \qquad\qquad (3)$$

Where is the seasonal length. For example, s = 12 for a monthly time series[26].

3.1 Analyzed Cases

The presented research analyzes data connected to top five crimes in Chicago. They are theft, assault, battery criminal damage, and narcotics. The Chicago crimes are punished by the section and act stated in Illinois Compiled statutes (ILCS).The ILCS database manipulation and updating is a continuous process. This paper also analyzes domestic violence and the arrest.

A.   Arrests

Arrests are known as detaining a person in custody by authority of law or seize someone by legal authority and takehim or her into custody. There are six types of arrest are there. They are Juvenile arrest, Felony arrest,hit, and run arrest, Private person's arrest, Misdemeanor arrest, and Warrant arrest. After arrest, police for questioning will take the arrested person to police station.

B.   Domestic Violence

Domestic violence is an aggressive of violent behavior of any people in home. Domestic violence is not physical violence alone but also includes mental torcher. Domestic violence is a kind of cruel behavior that forces control over the spouse, partner, girlfriend/boyfriend, or any close family member. Abusing a human is a learned behavior and is not caused by anger, drugs or alcohol, mental problem or common excuses.Division on Domestic Violence (DDV) handles the domestic violence crime.

C.   Top five crimes

As seen in the figure 1 it is understood that the top five crimes are theft, assault, battery, criminal damage and narcotics. The official crime code stated in Illinois Compiled Statutes is also included below:

a)   Theft and Assault

Theft is possession of other person's property or any service without any consent fromowner with the aim to bereave the property. Some of the theft are Retail Theft (720 ILCS 5/16-25), Identity Theft (720 ILCS 5/16-30),Theft or Possession of Stolen Vehicle (720 ILCS 5/19-2) and Online Theft. There are different methods available to commit theft, and the sanction for such crimes will be based on the seriousness and worth of item taken. The punishment can be jail or fines. An assault is defined as an act of causing unwanted physical harm or physical contact on a person by threat or any attempt to

threatening(720 ILCS 5/12-1). In the State of Illinois, if a person is found committing assault having different range of penalties like fine or significant prison time depending on the assault type. There are many factors are considered to decide the penalties. It includes whether the accused person using something that can be perceived as a weapon.
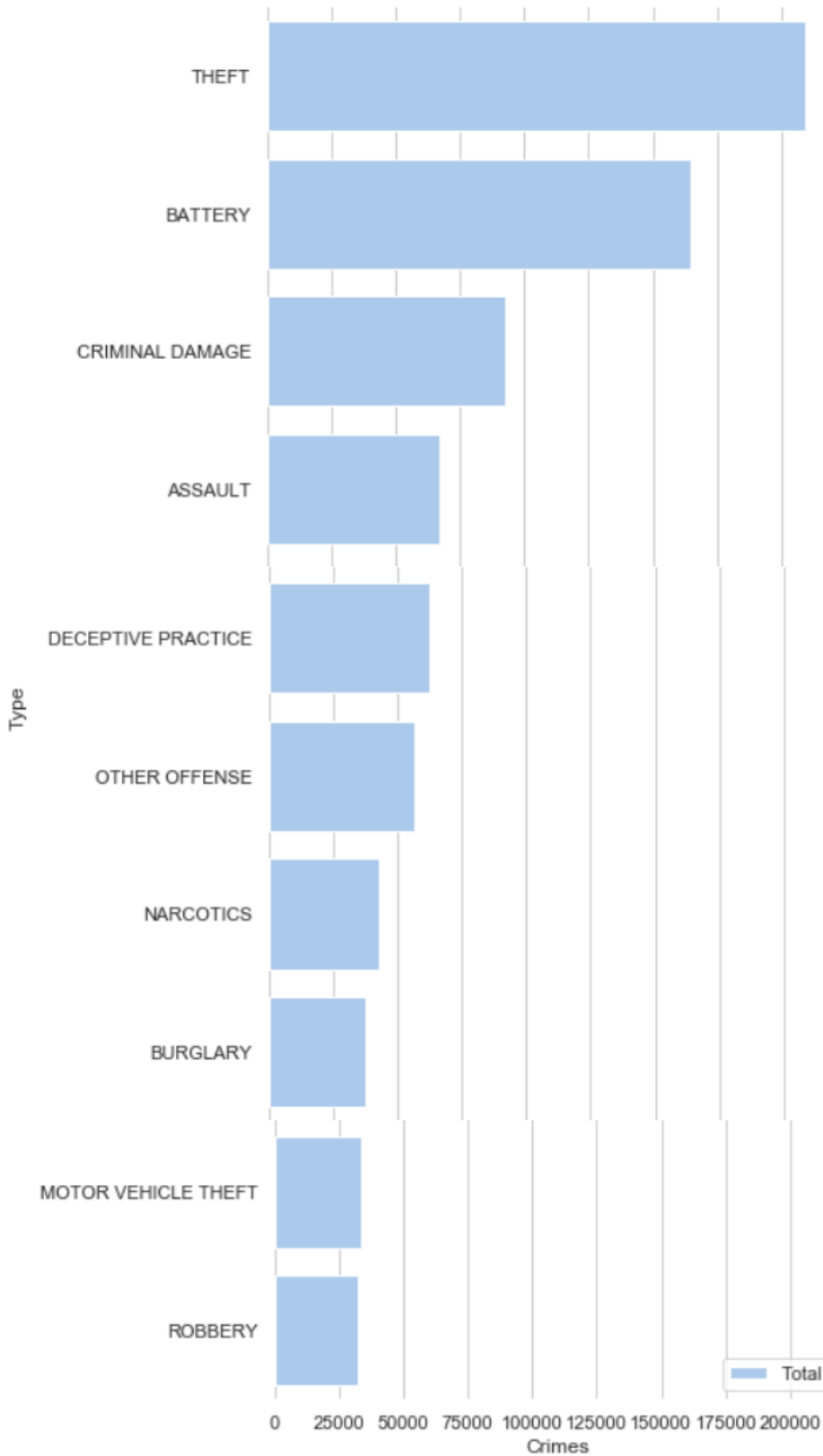

Figure 1. The top crimes of Chicago city.

b) *Battery and Criminal Damage*

Battery is the intentional, offensive and harmful contact with any another person. In criminal law the elements of battery are physical contact that causes harmful contact without that person's consent. Some of the elements that should be proven to establish a case for battery are as follows: (1) an act by a litigant (2) an intention to harm from the litigant and (3) harmful or offensive contact to the complainant. The act must result in one of three forms of contact. Pursuant to 720 ILCS 5/123, simple battery crime happens when one person causes physical harm to another with the intention of insulting, injuring or to provoke any person.Criminal damage is concentrated on the safety of dwellings and food supply and few other crimes were used for ruining any person's property without the permission from the owner. The charges that an accuse person have to face will be based on the type of damage and manner in which the damage has done. The offence range starts from Class 1 misdemeanor to class 2 felony, according to Illinois law (720 ILCS 5/21-1).

c) *Deceptive Practice*

A deceptive practice is a term that has often used in insurance claims cases to refer to instances in which an insurer acts in such a way as knowingly deceive the insured person about his or her coverage. Such crimes are usually white-collar offenses which intent for financial gain and most of the times they do not involve violence. Deceptive practice take place when he or she has the intention to defraud another person and might do (i) threatening a person to rob a document that may be a property or incurs a pecuniary obligation, (ii) misusing of position as an employee of any financial institution to create fraud in deposit, receipt or investment without the knowledge of institute and (iii) Knowingly makes another person to make public deceptive statement for the promotion to increase any kind of property sale or service sale. State legislation covers deceptive practice and fraud under 720 ILCS 5/17. Convictions can take place for compensation for the damage and the prison duration can be reduced if the victims are satisfied with the compensation amount.

4. Experimental Results

The time series analysis of crimes has done for the crime events in Chicago city which includes about 8 lakhs crime events. The main goal of this research is to predict the crime events based on crime took place in past and to plan the activities for the prevention of crimes. Crime data has been gathered from the Chicago Police Department website. The data is extracted from Chicago Police Department's CLEAR system. The dataset displays reported number of crimes except the murder in which each data involves a single victim.

The accuracy measures of the existing Arima algorithm and the proposed Arima algorithm shows that the proposed algorithm has more accuracy of 91.4%. The error rate has been measured Mean Absolute Error (MAE), Mean Square Error (MSE) and Root of Mean Square Error (RMSE). Figure 2 and Table 1 shows the accuracy measures and figure 3 and table 2 shows the error rates.
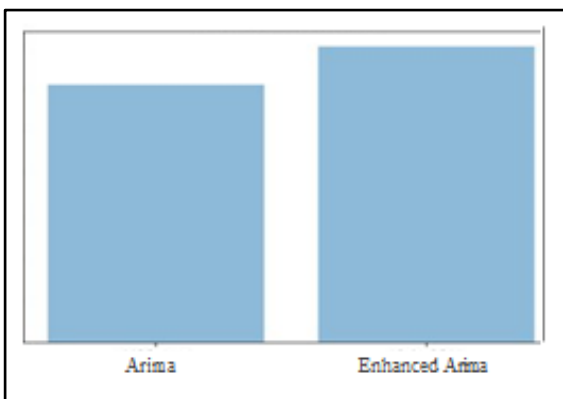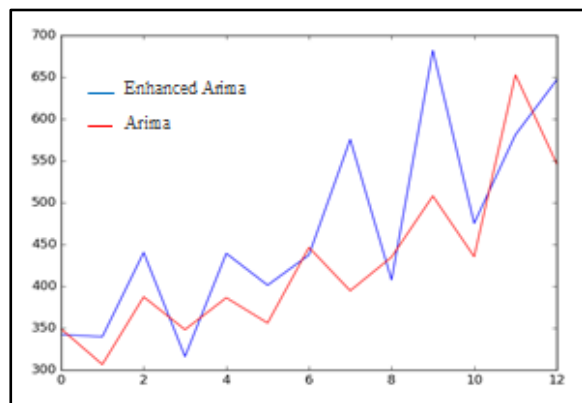

**Figure 2 Accuracy Measures**

Figure 3 Error Rates

Table 1 Accuracy Measures

Time Series Analysis Algorithms	Accuracy (%)
Existing ARIMA Algorithm	88.3
Proposed ARIMA Algorithm	91.4

Table 2 Error Rates

| Error Rates	Arima Model (%)	Enhanced Arima Model (%)
Mean Absolute Error (MAE)	11.7	8.6
Mean Square Error (MSE)	10.8	7.9
Root of Mean Square Error (RMSE)	10.2	7.2

The analyzed data included daily statements of crimes and offenses during the period from January 1, 2017 to May 27, 2020 of Chicago city. The data records contained 8,74,812 instances and 13 attributes namely ID, Case Number, Date, Block, Primary Type, Description, Location Description, Arrest, Domestic, Year, Updated On, Latitude and Longitude. The data was then aggregated according to the needs of the prediction model. The study presents the results of seven events such as arrest, domestic violence, theft, assault, battery, criminal damage anddeceptive practices. The data presented in the charts for hiding the actual crime rate. In figure 4 and figure 5. arrest plus domestic violence has been presented.

(a)   Yearly Arrest

(b) Monthly Arrest

(c) Weekly Arrest
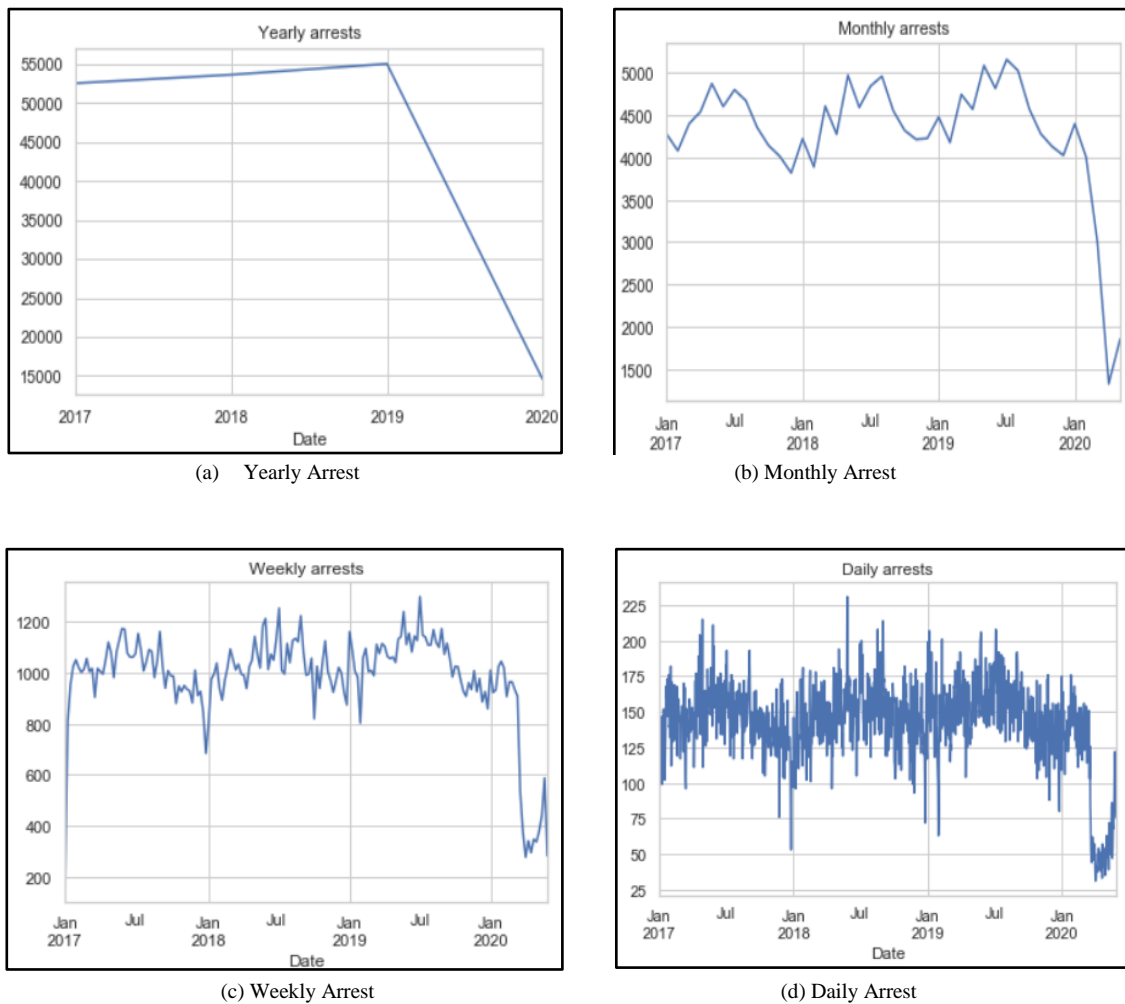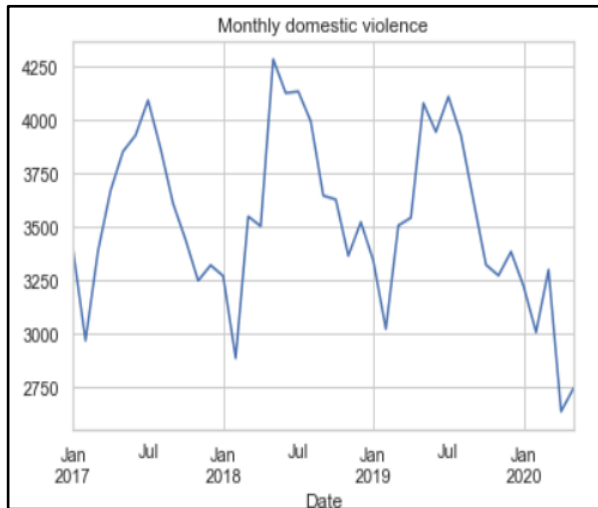
(d) Daily Arrest

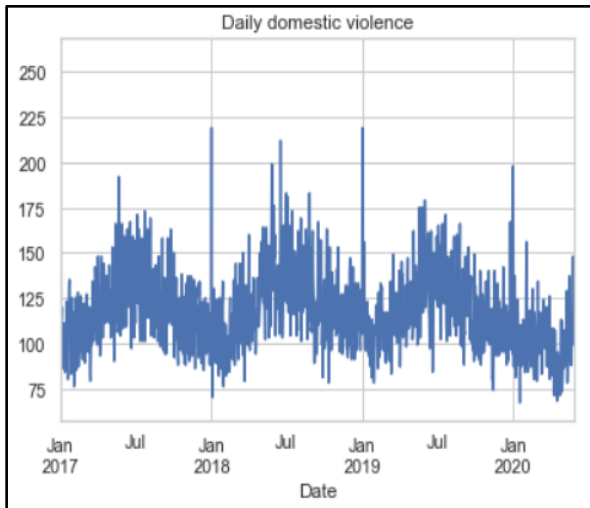**Figure 4 Represents Arrest of 2020 -2017**
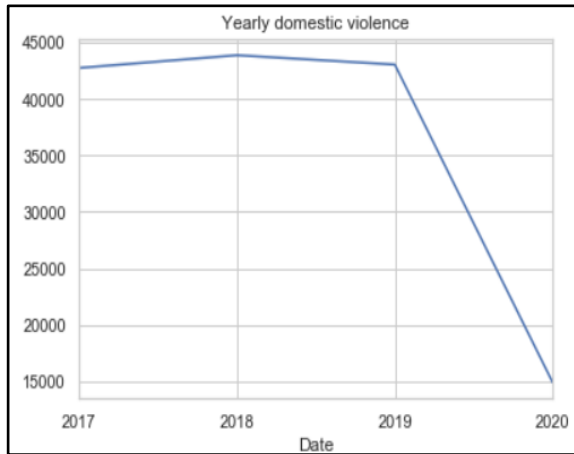
(a) Yearly Domestic Violence
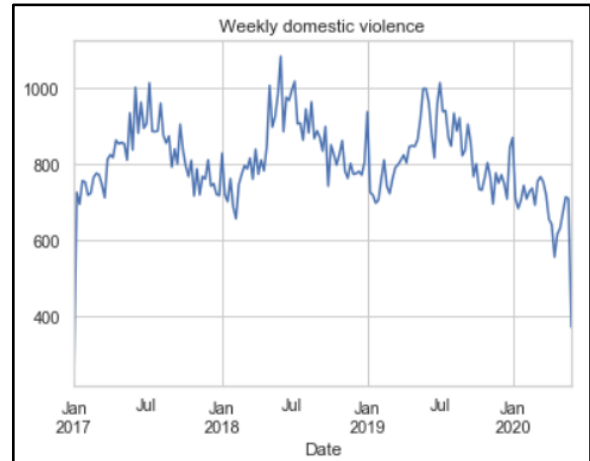

(b) Weekly Domestic Violence


(b)   Monthly Domestic Violence

(d) Daily Domestic Violence

Figure 5 Represents Domestic Violence of 2020 -2017

The ten crime types are analyzed and categorized based on counts of the crimes. The ten crime types which analyzed are theft, Battery, criminal damage, assault, deceptive practice, narcotics, burglary, motor vehicle theft, robbery and other offenses. Figure 1 shows crimes in Chicago city. From the figure, it is clear that top five crimes are theft, Battery, criminal damage, assault and deceptive practice respectively. Figure 6 shows the monthly seasonality of top five crimes. Figure 7 and figure 8 represents the weekly and daily seasonality of top five crimes.
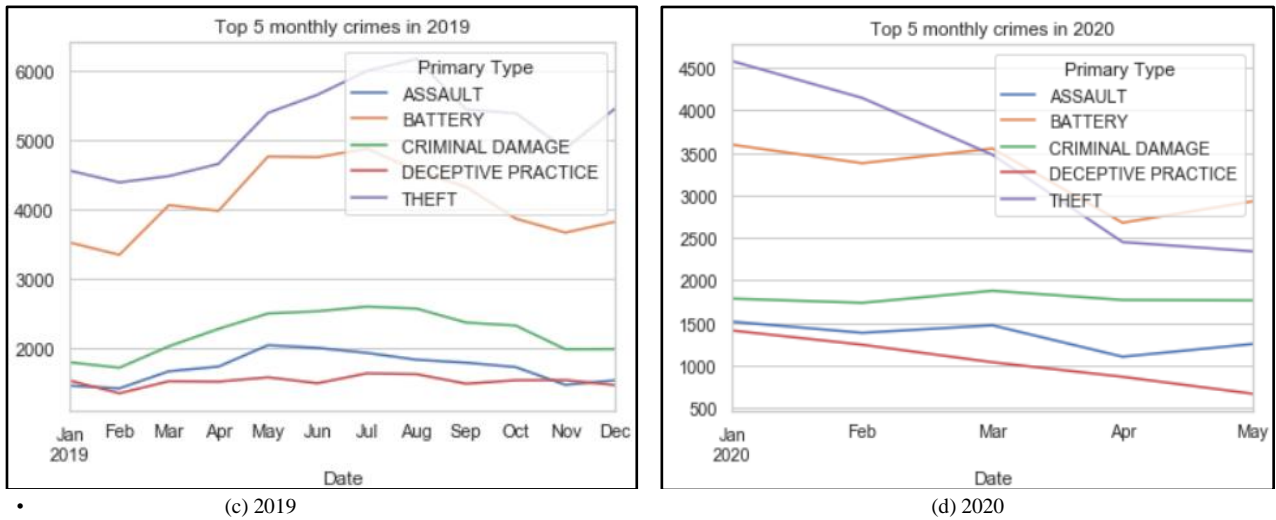

(a) 2017

(b) 2018

(c) 2019

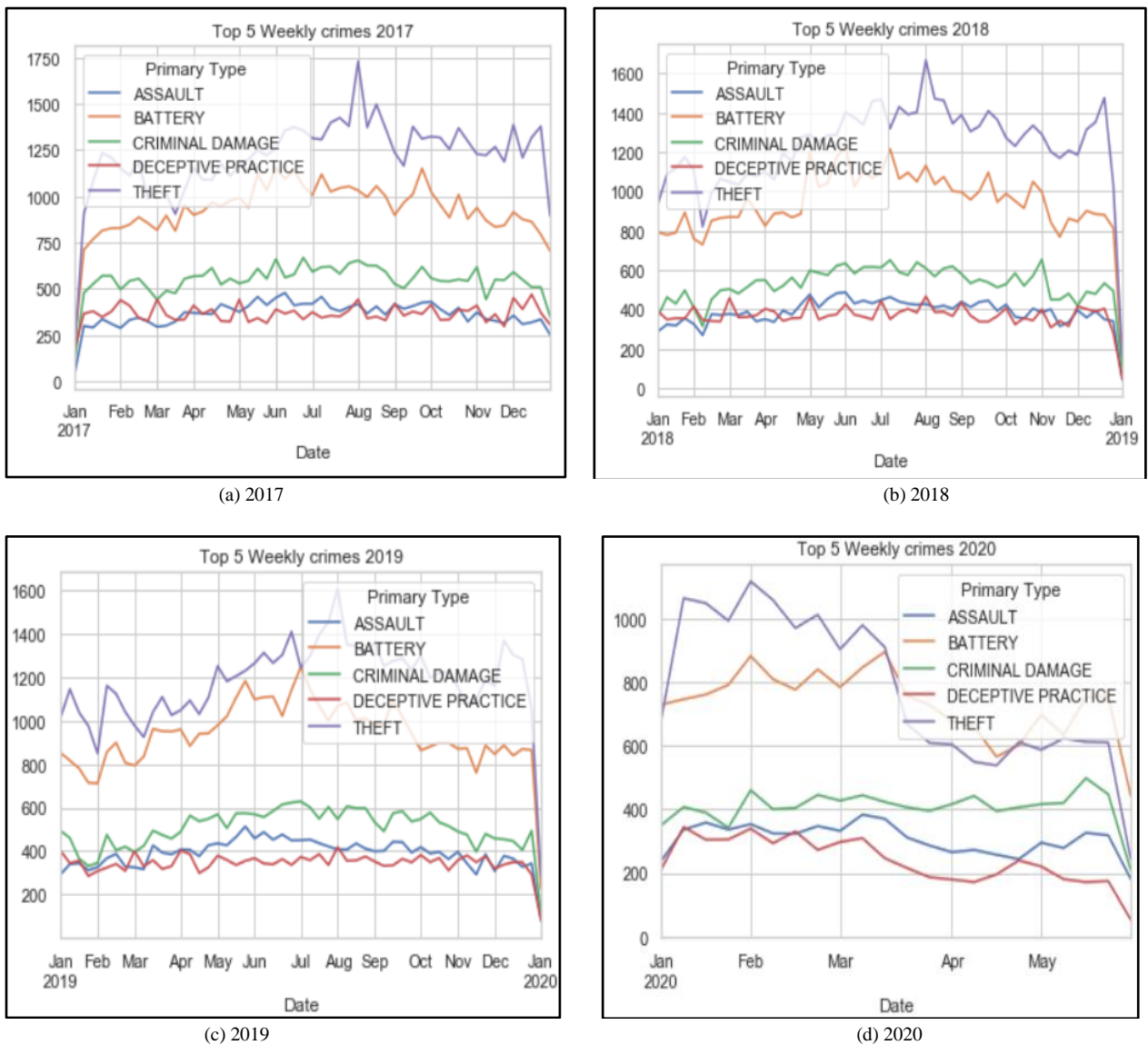(d) 2020
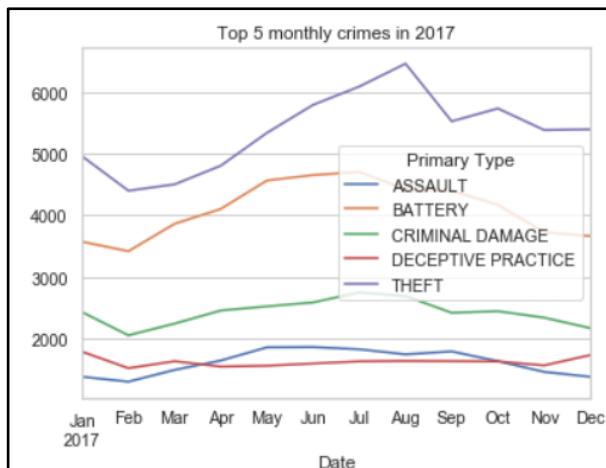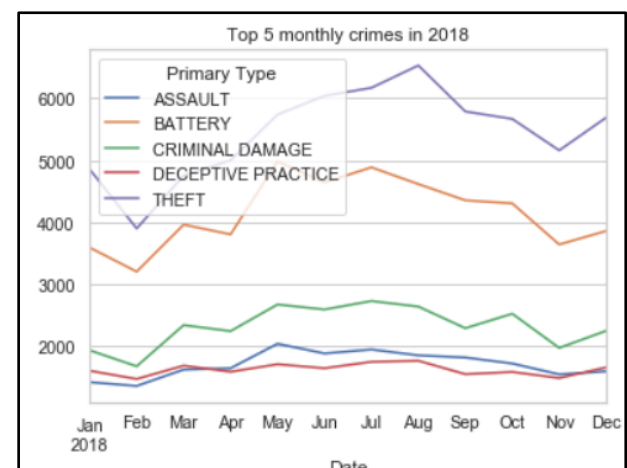
Figure 6 Represents the Top Five Monthly Crimes

(a) 2017

(b) 2018

(c) 2019

(d) 2020

Figure 7 Represents the Top Five Weekly Crimes
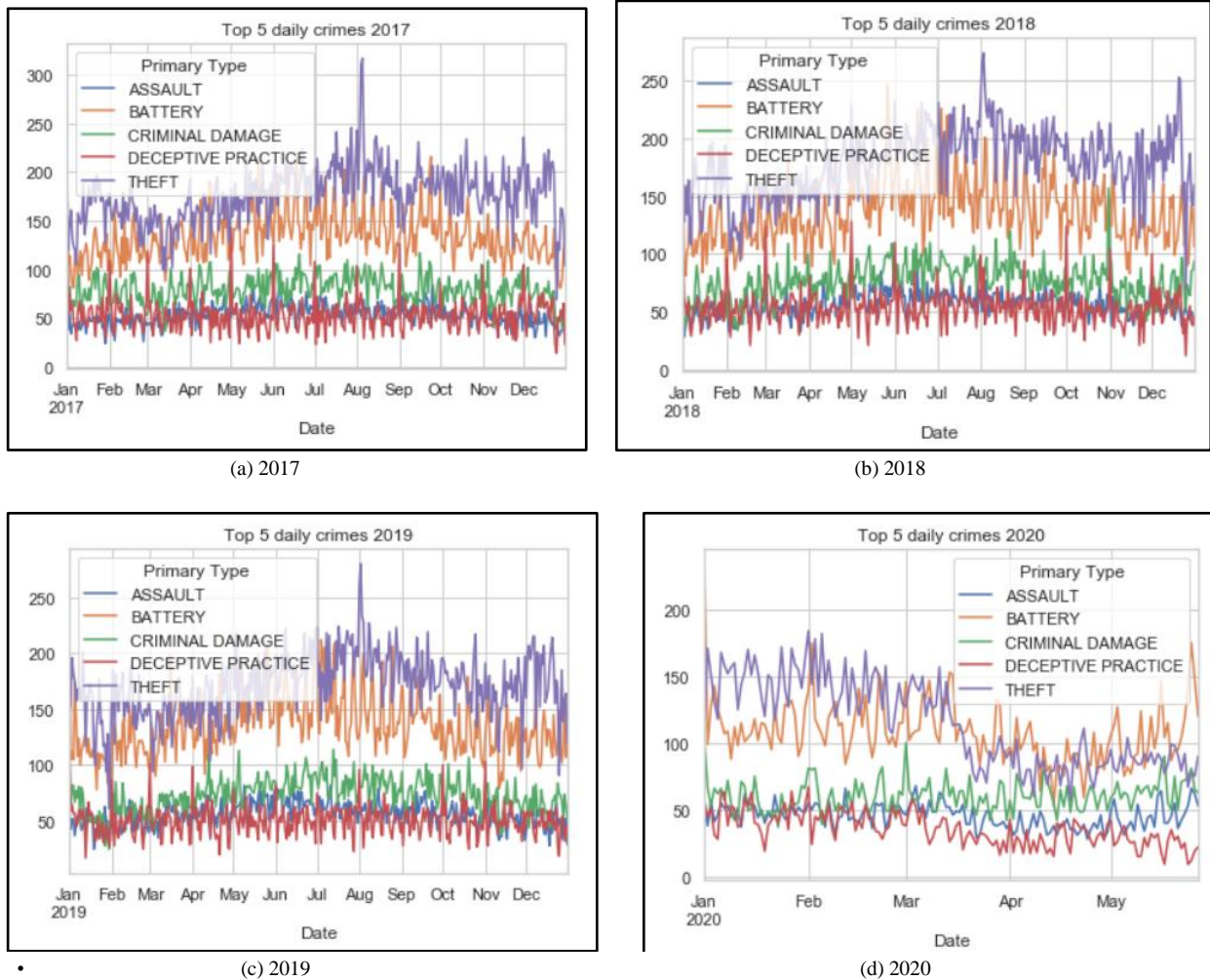
(a) 2017

(b) 2018

(c) 2019

(d) 2020

Figure 8 Represents the Top Five Daily Crimes

5. 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. This research is carried out using the Chicago crime data. From the beginning of 20th century, the Chicago crime incidents are tracked by the Chicago's Police Department's Bureau of Records. The dataset contains crimes from January 1, 2017 to May 27, 2020. Data mining is applicable in the context of law enforcement and crime related problem for crime prediction. The experimental results obtained by applying ARIMA model of time series of actual crime offense counts in Chicago confirm the utility of crime time series analysis. Monthly, Weekly and Daily seasonal patterns were identified. It is clear from the results that most of the crime took place in the duration May to August. The Chicago Police Department can make use of this prediction to take necessary action before the crime can occur. The result Future enhancement of this research work is to apply time series models to increase crime prediction accuracy and to enhance the overall performance. In addition, machine learning and optimization algorithms will be consider for better accuracy.

REFERENCES

[1] Sudan Jha, Eunmok Yang, Alaa Omran Almagrabi, Ali Kashif Bashir, Gyanendra Prasad Joshi "Comparative analysis of time series model and machine testing systems for crime forecasting", Springer-Verlag London Ltd., 30 January 2020.

[2] Md Abu Saleh, Ihtiram Raza Khan, "Crime Data Analysis in Python using K - Means Clustering", International Journal for Research in Applied Science & Engineering Technology (IJRASET), Volume 7 Issue IV, Apr 2019.

[3] Ayisheshim Almaw, Kalyani Kadam, "Survey Paper on Crime Prediction using Ensemble Approach", International Journal of Pure and Applied Mathematics, Volume 118, No. 8, 2018.

[4] Deepika K.K, Smitha Vinod, "Crime analysis in India using data mining techniques", International Journal of Engineering & Technology, 7, 2.6, 2018.

[5] Grzegorz Borowik, Zbigniew M. Wawrzyniak, Paweł Cichosz, "Time series analysis for crime forecasting", European Union, 2018.

[6] Dr.M.Sreedevi, A.Harsha Vardhan Reddy, Ch.Venakata Sai 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.

[7] Gourav Govindaswamy, 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.

[8] Md. Abdul Awal, Jakaria Rabbi, Imran Rana, "Using Data Mining Technique to Analyze Crime of Bangladesh", International Journal of Computer Science and Network, Volume 6, Issue 4, August 2017.

[9] Prajakta Yerpude, Vaishnavi Gudur, "Predictive Modelling Of Crime Dataset Using Data Mining", International Journal of Data Mining & Knowledge Management Process (IJDKP) Vol.7, No.4, July 2017.

[10] Sarpreet Kaur, Dr. Williamjeet Singh, "Systematic Review of Crime Data Mining", International Journal of Advanced Research in Computer Science, Volume 8, No. 5, May-June 2017.

[11] H. Benjamin Fredrick David, A. Suruliandi, "Survey On Crime Analysis And Prediction Using Data Mining Techniques", Ictact Journal On Soft Computing, Volume: 07, Issue: 03, April 2017.

[12] Arnab Samanta, Amol Joglekar, "Crime Classification and Criminal Psychology Analysis using Data Mining", International Journal of Mechanical Engineering and Information Technology(IJMEIT), Volume 4, Issue 10, October 2016.

[13] Tushar Sonawanev, Shirin Shaikh, Shaista Shaikh, Rahul Shinde, Asif Sayyad, "Crime Pattern Analysis, Visualization And Prediction Using Data Mining", IJARIIE, Vol-1, Issue-4, 2015.

[14]Lenin Mookiah,William Eberle, Ambareen Sira, "Survey of Crime Analysis and Prediction", Twenty-Eighth International Florida Artificial Intelligence Research Society Conference, 2015.

[15] Tahani Almanie, Rsha Mirza, Elizabeth Lor, "Crime Prediction Based On Crime Types And Using Spatial And Temporal Criminal Hotspots", International Journal of Data Mining & Knowledge Management Process (IJDKP) Vol.5, No.4, July 2015.

[16] Shiju Sathyadevan, Devan M.S, Surya Gangadharan. S, "Crime Analysis and Prediction Using Data Mining", Conference Paper, August 2014 DOI: 10.1109.

[17] Dr: Zakaria Suliman Zubi, Ayman Altaher Mahmmud, "Using Data Mining Techniques to Analyze Crime patterns in the Libyan National Crime Data", Conference Paper, December 2013.

[18] Dawei Wang, Wei Ding, Henry Lo, Tomasz Stepinski, Josue Salazar, Melissa Morabito, "Crime hotspot mapping using the crime related factors—a spatial data mining approach", Springer Science+Business Media New York 2012, December 2012.

[19] 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(I JCST), Vol. 4, ISSue 2, AprIl - June 2013.

[20] Suncica Vuji´c, Jacques Commandeur, Siem Jan Koopman, "Structural Intervention Time Series Analysis of Crime Rates: The Impact of Sentence Reform in Virginia", Tinbergen Institute Discussion Paper, January 30, 2012.

[21] Quanbao Jiang, Jesús Javier Sánchez Barricarte, "A Crime Rate Forecast and Decomposition Method", International Journal of Criminology and Sociological Theory, Vol. 4, No. 2, December 2011.

[22] Fathima Afroz, Rajashekara Murthy S , Chayadevi M.L, "Crime Analysis and Prediction Using Data Mining – CAP a Survey".

[23] Arwa Abu Shmais, Rana Hani, "Data Mining for Fraud Detection", Prince Sultan University, Saudi Arabia.

[24] Peng Chen, Hongyong Yuan, Xueming Shu, "Forecasting Crime Using the ARIMA Model", Fifth International Conference on Fuzzy Systems and Knowledge Discovery, IEEE, 2008.

[25] Maryam Farsi, Alireza Daneshkhah, Amin Hosseinian-Far, Omid Chatrabgoun, and Reza Montasari, "Crime Data Mining, Threat Analysis and Prediction", Springer Nature Switzerland, Advanced Sciences and Technologies for Security Applications, 2018.

[26] Sanjeev Sridharan, Suncica Vuji´c, Siem Jan Koopman, "Intervention time series analysis of crime rates: the impact of sentence reforms in Virginia", Tinbergen Institute Discussion Paper, TI 2003-040/4