

## International Journal of Computer Science and Mobile Computing



A Monthly Journal of Computer Science and Information Technology

ISSN 2320-088X

IMPACT FACTOR: 6.199

*IJCSMC, Vol. 8, Issue. 6, June 2019, pg.134 – 143*

# Crime Analysis for Multistate Network using Naive Bayes Classifier

**Mrinalini Jangra**

Student, M-TECH CSE

Dav Institute of Engineering and Technology

Jalandhar, Punjab 144022, India

[jangrameenu321@gmail.com](mailto:jangrameenu321@gmail.com)

**Ms. Shaveta Kalsi**

Assistant Professor, CSE

Dav Institute of Engineering and Technology

Jalandhar, Punjab 144022, India

[shaveta\\_gne@yahoo.com](mailto:shaveta_gne@yahoo.com)

**ABSTRACT:** *The process which is used to extract all necessary and useful information for data analysis is called data mining. The K-nearest neighbor classifier is utilized to compute good performance optimal values. Bayesian Network is a graphical model. This model is utilized to establish associations which are beneficial for a set of variables. These networks represent Statistical learning algorithms. The associations that are structural in behavior occur for old information. These classifiers cannot be applied for data sets having large number of features and this is the main drawback of this classifier. The performance of Naïve Bayes classifier is compared with the KNN classifier. The proposed approach is applied in Anaconda. The simulations results depicts that Naïve Bayes algorithm has high accuracy rate and less execution time.*

**Keywords:** *Crime Prediction, Naïve Bayes, KNN, Prediction Analysis*

## 1. INTRODUCTION

In prediction analysis, future tendencies and results are predicted on the basis of supposition. The prediction analysis approach uses machine learning methods and regression mechanisms for conducting predictive analytics. The machine learning methods are utilized in huge manner to perform predictive analysis. These techniques have become trendy because these techniques are capable to handle large scale information effectively and also show good performance. These techniques give results with uniform features and noisy information [1]. The pioneering predictive models have been applied in numerous fields just like health care, cyber safety, learning, credit card scam discovery, social media, cloud computing, software extent, excellence and imperfection forecasting, price and attempt opinion software reuse and so on. The dependent and independent variables and their association can be analysed using prediction technique [2]. The technique can be used in future to get improvised results.

The Data mining approach is utilized in almost all kind of applications. The process builds representative model in which observational data is implemented. The prediction model serves two purposes. Firstly, the output is predicted on the basis of the input variables and secondly, this model helps in understanding the relationship between the output variable and all input variables.

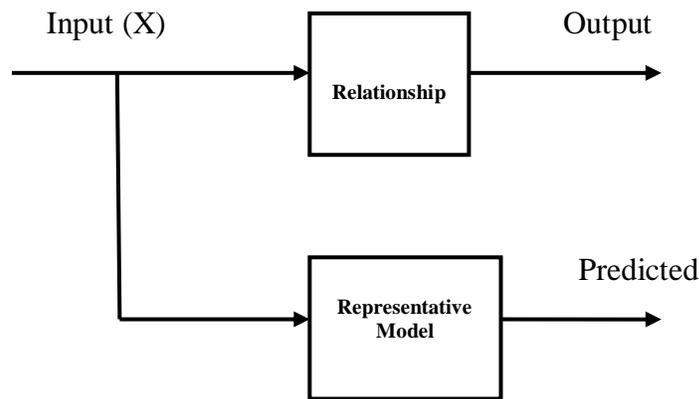


Fig 1.1: Prediction analysis model

Crime is a part of society which needs to be recovered as there is rule of law. In this act, all rules abided by the law are forbidden or omitted by the criminals for which punishments are imposed which is a long term procedure [3]. The crimes cannot be predicted easily. In this act, the criminal does not obey law and make a miss happening that affects the secure life of society. Therefore, there is a need of appropriate or good theoretical understanding so that practical solutions can be provided for the prevention of crime in which all details of time and location is provided correctly [4]. The crime analysis uses past crime data for predicting future crime locations and time. Crime prediction for future crime is a procedure that finds out the rate of change of crime in every year and projects those changes for future. This process utilizes both qualitative and quantitative methods. Environmental scanning, scenario writing are some attributes using which future nature of criminal activity can be predicted. Therefore, crime measurable methods have been utilized to predict the future scope of crime. The projection of annual crime rate trends developed through time series models is a common method developed for crime prediction.

In the crime analysis, crime data is utilized through which law is enforced for capturing criminals and prevention of crimes [5]. The patterns are extracted manually from past centuries from the data. An improvement in data collection, storage and manipulations can be seen due to the propagation, universality and advancement of computer technology. The data mining process is used to implement these methods are applied through data mining procedure for revealing hidden patterns. Data mining techniques have been used to analyze and observe people's behavior [6]. Thus, in this procedure hidden patterns are revealed and novel patterns are identified from large data sets.

The aim of prediction technique is to forecast the values of attributes on the basis of other attribute's values. The first model is created on the basis of distribution of data. This model is utilized to predict future on unknown value [7]. The fundamental techniques of data mining are commenced for predicting crimes.

**a. Naive Bayesian (NB) Networks:** In this type of network, acyclic graphs exist with parent and with several numbers of children. In this network, it is assumed that child nodes are independent of parent nodes. The Bayes classifiers are less accurate in comparison with other classifiers. In some situations, it has been found that naive Bayes is better than few other classifiers.

**b. Multi-layer Perceptron:** In this classifier, the problem related to linear constraint is solved to attain the weight of network [19]. Several other well known algorithms work on the idea of perceptron. The continuous running of a training batch is a process in which this algorithm is utilized.

**c. K-means:** This algorithm is most simple algorithm among all other algorithms. This is an unsupervised learning algorithm. Large number of data is clustered using easy and efficient manners. This algorithm is utilized in the

absence of labeled data. Therefore, the reliability of this algorithm relies on distance function. The Euclidean distance method is applied to compute minimum average distance. The K-nearest neighbor classifier is utilized to compute good performance optimal values.

**d. Decision Trees:** The instances are classified on the basis of arrangement of feature values. Each node represents a feature of an instance while the implicit value is offered in the form of a branch [7]. This algorithm is used in various machine learning and data mining methodologies. In this model, an item is monitored to identify target value of item. Regression or classification trees are other names given to decision tree.

**e. Neural Networks:** In this network, entire multi-processing components are collected jointly with the help of adaptive non-linear data processing algorithms. Classification of these systems is performed on the basis of the natural- adapting and self-organization. The methods in the neural networks are simulated on the basis of the coding for the neurons in any creature. Various regression/ classification techniques can be implemented in these networks at once dissimilar from other systems in which only one technique is implemented at a time [7]. The performance of the Artificial Neural Network relies on the three rudiments like weight of each input connection, network architecture and activation functions. This technique is introduced for the prediction of offense and focuses on biological areas of apprehension that surpass conventional police force limits.

**f. Bayesian Network:** Bayesian Network is a graphical model. This model is utilized to establish associations which are beneficial for a set of variables. These networks represent Statistical learning algorithms. The associations that are structural in behavior occur for old information. These classifiers cannot be applied for data sets having large number of features and this is the main drawback of this classifier.

#### **g. Support Vector Machine**

Vladimir Vapnik and his colleagues proposed the Support Vector Machines. This method has gained popularity in recent times in the area of machine learning and pattern classification. This classification algorithm is based on optimization hypothesis [8]. This approach is also recognized as binary classifier because it exploits the margin. The best hyperplane separates all the information points of an individual class. This can be recognized throughout the classification offered by SVM. In the SVM, the best hyperplane is explained by the biggest periphery amid the two classes. The utmost margin in hyperplane is alienated with the help of SVM classifier. Thus, this method can prove extremely helpful and advantageous in crime forecasting.

**h. Fuzzy Time Series:** This technique provides results even some data is not available. This approach is used for the offense prediction. The earlier projected techniques are restricted to several extents because of which this method is applied for the prediction of crime and for offering practical computation methods. This method is applied in different data for checking its working whether it give effectual and proficient outcomes or not. However, attained outcomes were acceptable and accurate in comparison with other techniques in the forecasting of offense. This method gives outcomes even in the unavailability of some information.

## **2. LITERATURE SURVEY**

S Prabarakan, et.al (2018) stated that design and patterns from large dataset were discovered using data mining approach. This approach involved some processes for this purpose. In the convergence of machine learning and database structure, this approach included different policies [9]. The analysis of market basket, learning, manufacturing engineering, crime inquiry and future healthcare were some areas in which this approach could be implemented. The inquiry of crime was an appealing application which was used for the processing of offense features. The investigation of crime helped society for a better living. In this study, different data mining techniques utilized in this field were reviewed. In future, this review can provide support in the designing of novel approaches for the forecasting of offense and investigation.

Prajakta Yerpude, et.al (2017) stated that enhancement in crime rates was seen across the globe that required to be diminished. For this purpose, proper analysis was needed so that crime rate could be minimized to minimal. With the help of this analysis, it became easier for police to take suitable steps in solving crime as soon as possible [10].

Various data mining techniques related to crime data were implemented for predicting features through which high crime rates were affected. Data sets had been utilized for training, testing and getting desired results by using supervised learning. On the other hand, inconsistent, unstructured data had been divided by unsupervised learning into several classes or clusters. There were some supervised learning techniques in data mining and machine learning like decision trees, Naïve Bayes and Regression on earlier gathered data. This data was the reason of crime in a region or locality by which features could be predicted. Essential actions had been taken by the Crimes Record Bureau and Police Department on the basis of ranking of the features for decreasing the chances of offense incidence.

Shyam Varan Nath, *et.al* (2016) stated that the problems related to the recognition of crime could be removed using data mining. The entire society suffered due to crimes as it caused social annoyance. About 50% of the offenses were committed by 10% of the offenders. In order to design the proactive services using data mining technology which reduced crime activities could be controlled effortlessly and early. In the data mining approach, the clustering algorithm was utilized by which crimes patterns were identified and the process of crime solving increased. For providing help to process identification of crime patterns, k-means clustering was enhanced. All these techniques were implemented in the real time applications for validating all outcomes. In order to attain knowledge discovery from the crime records, semi-supervised learning technique had been utilized which increased prediction precision [11]. In this study, a weighting scheme for attributes was developed so that limitations of box clustering tools and techniques could be handled. The execution of a data mining structure which worked with geospatial crime plot was quite easy. It also enhanced the productivity of investigators and other law enforcement agencies. This approach could also be utilized for the implementation of activities against terrorism.

Dr. Nevine Makram Labib, *et.al* (2015) stated that data mining technique was extensively utilized in almost all applications and became a powerful tool in all domains. In the field of counter terrorism, these tools had been utilized extensively. In this study, data mining applications utilized in predictive policing were reviewed. The main aim of this study was to propose and suggest a data-mining model for the prediction of most important factors that influenced crime incidence [12]. Different data mining techniques such as decision trees, Naïve Bayes, and Association rules were utilized to compare the efficiency of each algorithm. Among all algorithms, best results were provided by the association rules. The tested results demonstrated that proposed model must be implemented as a system which could be used directly by the police from police head quarters for several levels of decision making.

Lawrence McClendon *et.al* (2015) stated that for the detection and prevention of crime, an essential role was played by both data mining and machine learning approaches. An open source data mining software called WEKA had been utilized in this study using which a comparative analysis of patterns of violent crimes was conducted from various communities and the dataset provided by the university of California regarding incidence of crime [13]. On the communities and crime dataset, the Linear Regression, Additive Regression, and Decision Stump algorithms were implemented through the usage of same finite set of features. On the basis of performed experiments, it was concluded that performance of the linear regression algorithm was better in comparison with other two algorithms. The effectiveness and accuracy of the machine learning algorithms used in data mining analysis was proved by means of which all aggressive crimes were forecasted.

Shiju Sathyadevan, *et.al* (2014) stated that study and avoidance of offense was a systematic approach using which patterns and trends were identified and analyzed properly. The proposed model was able to forecast high crime incidence possible areas. This model could also visualize crime vulnerable regions. The process of crime solving became speedy due to the expansion of automated systems. It was possible to extract the earlier unidentified and valuable information from an unstructured data with the help of data mining approach [14]. In this study, an approach was proposed for developing a data mining process for quick crime solving. This approach was applied among the computer science and criminal justice. In this study, the attention was given to the crime aspects of daily life. This approach did not consider the reasons of offense incidence such as scandalous background of criminal, political acrimony and so on.

Rasoul Kiani, et.al (2015) stated that organizations and institutions were affected by frequently occurred offenses someway. Therefore, it was imperative to analyze factors, causes and connection of different crimes and to find out suitable means for controlling and avoiding further offenses. The classification of clustered crimes on the basis of occurrence frequency during different years was the major aim of this study. In terms of scrutiny, research and detection of patterns for the incidence of several offenses, data mining techniques had been utilized [15]. In this study, the weights to the features were assigned by which quality of the model was improved and low value was removed. For the optimization of Outlier Detection operator parameters, the RapidMiner tool was utilized by Genetic Algorithm (GA).

Tahani Almanie, et.al (2015) stated that the identification of spatial and temporal criminal hotspots was the main objective of this study. Using this technique, it was also identified that how one could run Apriori algorithm which was used to generate interesting frequent patterns for the criminal hotspots. In this study, several classification models such as Decision Tree classifier and Naïve Bayesian classifier were used for assisting in the prediction of possible offense types. In this study, the analysis was reviewed in detail in which all obtained Denver crimes' dataset with its demographics information were utilized for the analysis of crime [16]. It was also utilized to capture the factors which influenced the safety of neighbors. The obtained results were utilized to aware people about the dangerous places and also predicted future crimes in a particular place within a specific time period that helped all agencies.

### 3. RESEARCH METHODOLOGY

This crime prediction in India is the base of this research work. The KNN classifier is applied in proposed work for prediction analysis. K-Nearest-Neighbor (kNN) is an effective and effective non-parametric classification approach that is utilized in several scenarios. In order to classify data record  $t$ , the  $k$ -nearest neighbors are extracted. The neighborhood of  $t$  is produced as well. Amongst the data records present in the neighborhood, the classification of  $t$  can be determined by means of majority of voting. The distance-based weighting may be either accessible or not. On the other hand, for the application of kNN, a suitable value of  $k$  is needed to be chosen. The success of classification just depends on this value. Therefore, the kNN approach is absolutely reliant on value of  $k$ . This selection of this value can be done in several ways. Though, one easy way is the running of algorithm several times with different values of  $k$ . In this approach, the best performance giving value is chosen. With the help of KNN classification model, novel examples can be classified at higher cost. Several applications avoid using this algorithm as it is a lazy learning technique, therefore, numerous applications do not use this algorithm. Identification of representatives to represent absolute training data for classification is one way that is use to improve the efficiency of this approach. The training dataset is utilized for the generation of an inductive learning model. Moreover, this model is utilized for classification purpose.

The technique of maximum likelihood is utilized in different practical applications through the estimation of parameter for Naïve Bayes classifiers.

Thus, a conditional model is identified as a probability model for a classifier

$$p(C|F_1, \dots, F_n)$$

This model comprises a dependent class variable  $C$  having less number of results or classes.  $F_1$  variables are several variables provided by  $F_n$ . It is infeasible to base such a model on probability tables in case of huge numbers of features or when a feature takes large numbers of values. Therefore, this model is again formulated in order to make this model more traceable. The equation generated using Bayes' theorem is given below:

$$p(C|F_1, \dots, F_n) = \frac{p(C)p(F_1, \dots, F_n|C)}{p(F_1, \dots, F_n)}$$

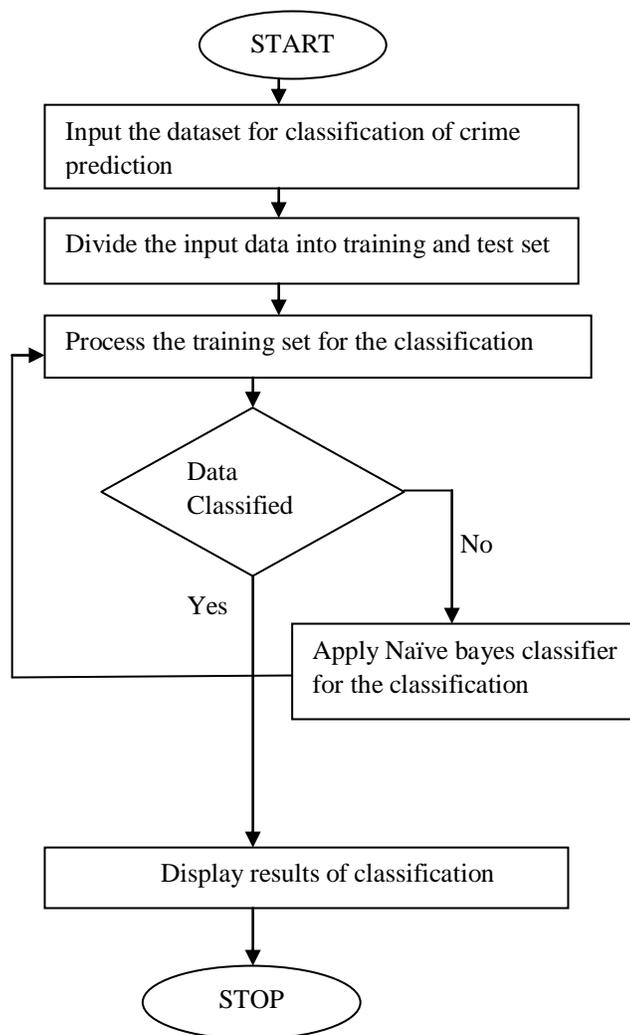


Fig 3.1: Proposed Methodology

#### 4. RESULT AND DISCUSSION

Python is an objective oriented language. This language includes semantics in dynamic form. This language is used to develop application rapidly. High level data structures along with the features of dynamic typing and binding make this language highly popular. This language is enormously utilized for scripting purpose for getting accessible components together. The syntax of python is simple. This learning process of this language is extremely simple which reduces maintenance cost. Python language assists modules and packages and this encourages the modularity of programs and reutilization of existing code.



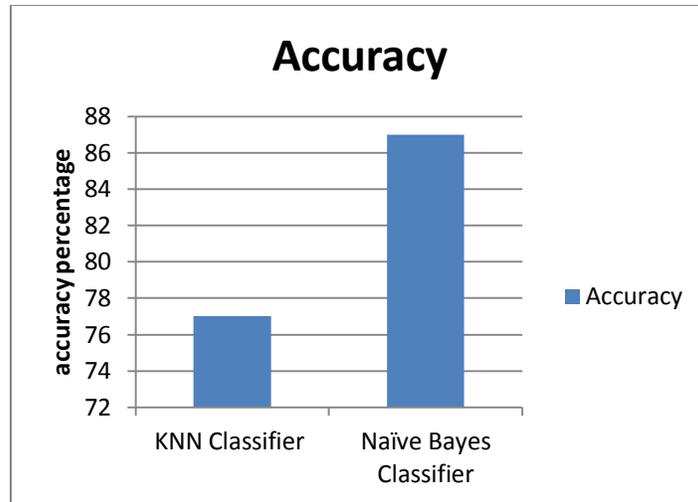


Fig 4: Accuracy Comparison

The performance of KNN and Naïve Bayes classifier is analyzed in terms of accuracy as depicted by figure3.3. The Naïve Bayes classifier shows better accuracy in comparison with KNN classification model.

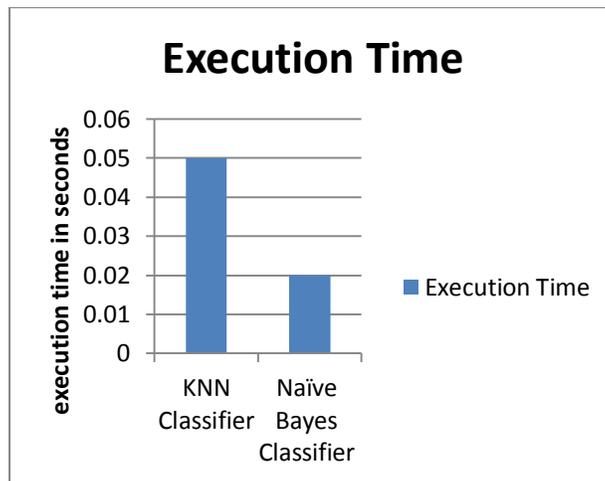


Fig 5: Execution Time

The performance of KNN and Naïve Bayes classifier is analyzed in terms of execution time as depicted by figure 5. The Naïve Bayes classifier requires less execution time in comparison with KNN classification model.

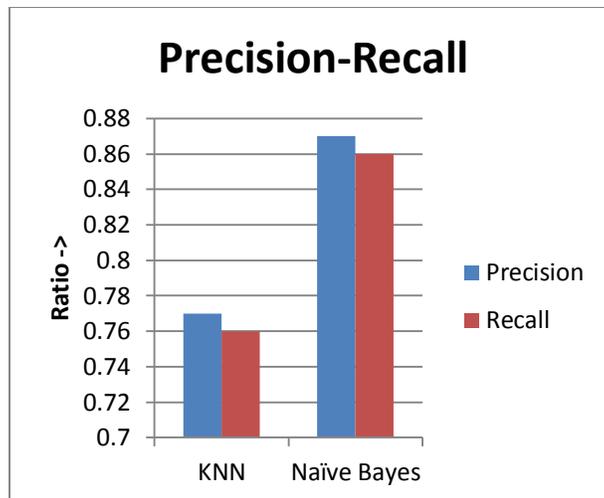


Fig 6: Precision-Recall Analysis

The performance of KNN and Naïve Bayes classifier is compared in terms of precision-recall. The value of Naïve Bayes classifier is high as compared to KNN in terms of precision and recall.

## 5. CONCLUSION

For developing information about reason, timing and location of crime occurrence, an analysis is performed. In this process, data is analyzed for the prediction of crime. The main aim of this analysis is to provide support to police officials and detectives to identify and understand particular and immediate crime issues. This analysis works together with patrolling officers and detectives. The existing patterns related to the activities of offenders are identified for predicting all possible future crime events as a main aim of this analysis. In this research work, a novel crime prediction technique is implemented. This approach is based on KNN and naïve bayes technique. The proposed and earlier approaches of crime prediction are applied in anaconda. Accuracy and execution time are utilized to analyze prediction results. This research work implements Naïve bayes classifier for the prediction of offense.

## References

- [1] Arthisree K.S and Jaganraj A, "Identify Crime Detection Using Data Mining Techniques", ISSN: 2277 128X, IJARCSSE, 2013.
- [2] P.Charanya, T.Dhivyabharathi, B.Kothai, J.Nivetha, "Forge Detection in Credit Application", ISSN: 2320-2106, International Journal of Advanced Computational Engineering and Networking.
- [3] Shyam Varan Nath, "Crime Pattern Detection Using Data Mining", WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology, IEEE, 2016.
- [4] Ubon Thongsatapornwatana, "A Survey of Data Mining Techniques for Analyzing Crime Patterns", ISSN: 978-1-5090-2258- 8/16, IEEE, 2016.
- [5] Isuru Jayaweera, Chamath Sajeewa, Sampath Liyanage, Tharindu Wijewardane, Indika Perera, Adeesha Wijayasiri, "Crime Analytics: Analysis of Crimes through Newspaper Articles", ISSN: 978-1-4799- 1740-2/15, IEEE, 2015.
- [6] Mugdha Sharma, "Z - CRIME: A Data Mining Tool for the Detection of Suspicious Criminal Activities Based on Decision", ISSN: 978-1-4799-4674-7/14, IEEE, 2014.
- [7] Li Ding, Dana Steil, Matthew Hudnall, Brandon Dixon, Randy Smith, David Brown, Allen Parrish, "PerpSearch: An Integrated Crime Detection System", ISSN: 978-1-4244-4173-0/09, IEEE, 2009.

- [8] Devendra Kumar Tayal, Arti Jain, Surbhi Arora , Surbhi Agarwal , Tushar Gupta , Nikhil Tyagi, "Crime detection and criminal identification in India using data mining techniques", *AI & Soc* (2015) 30:117– 127, Springer-Verlag London 2014.
- [9] S Prabakaran and Shilpa Mitra, "Survey of Analysis of Crime Detection Techniques Using Data Mining and Machine Learning", *National Conference on Mathematical Techniques and its Applications (NCMTA 18)*, 2018.
- [10] Prajakta Yerpude1 and Vaishnavi Gudur, "Predictive Modeling of Crime Dataset Using Data Mining", *International Journal of Data Mining & Knowledge Management Process (IJDKP)* Vol.7, No.4, July 2017.
- [11] Shyam Varan Nath, "Crime Pattern Detection Using Data Mining", *IEEE*, 2016.
- [12] Dr. Nevine Makram Labib, Brigadier-General Wael Kamal Arafa, "A Proposed Data Mining Model to Enhance Counter- Criminal Systems with Application on National Security Crimes", *International Research Journal of Computer Science (IRJCS)* ISSN: 2393-9842 Issue 7, Volume 2 (July 2015) .
- [13] Lawrence McClendon and Natarajan Meghanathan, "Using Machine Learning Algorithms To Analyze Crime Data", *Machine Learning and Applications: An International Journal (MLAIJ)* Vol.2, No.1, March 2015.
- [14] Shiju Sathyadevan, Devan M.S, Surya Gangadharan. S, "Crime Analysis and Prediction Using Data Mining", *IEEE*, 2014.
- [15] Rasoul Kiani, Siamak Mahdavi, Amin Keshavarzi, "Analysis and Prediction of Crimes by Clustering and Classification", *International Journal of Advanced Research in Artificial Intelligence*, Vol. 4, No.8, 2015.
- [16] Tahani Almanie, Rsha Mirza and 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.