IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

CRIME ANALYSIS AND PREDICTION USING DATAMINING: A REVIEW

1Shanjana A.S, 2Dr.R.Porkodi

1PGStudent(Department of computer Science), 2Associate Professer(Department of Computer Science)

1Bharathiar University,

2Bharathiar University

ABSTRACT

Crime analysis and prediction is a systematic approach for identifying the crime. This system can predict region which have high probability for crime occurrences and visualize crime prone area. Using the concept of data mining we can extract previously unknown, useful information from an unstructured data. The extraction of new information is predicted using the existing datasets. Crimes are treacherous and common social problem faced worldwide. Crimes affect the quality of life ,economic growth and reputation of nation.

With the aim of securing the society from crimes, there is a need for advanced systems and new approaches for improving the crime analytics for protecting their communities. We propose a system which can analysis, detect, and predict various crime probability in given region. This paper explains various types of criminal analysis and crime prediction using several data mining techniques.

KEYWORDS

Crime prediction, Decision trees, Linear Regression, k-means.

1. INTRODUCTION

Day by day crime data rate is increasing because the modern technologies and hi-tech methods are helps the criminals to achieving the illegal activities .according to Crime Record Bureau crimes like burglary, arson etc have been increased while crimes like murder, sex, abuse, gang rap etc have been increased [2].crime data will be collected from various blogs, news and websites. The huge data is used as a record for creating a crime report database. The knowledge which is acquired from the data mining techniques will help in reducing crimes as it helps in finding the culprits faster and also the areas that are most affected by crime

[11].Data mining helps in solving the crimes faster and this technique gives good results when applied on crime dataset, the information obtained from the data mining techniques can help the police department.

A particular approach has been found to be useful by the police, which is the identification of crime 'hot spots 'which indicates areas with a high concentration of crime [1].Use of data mining techniques can produce important results from crime report datasets. The very step in study of crime is crime analysis. Crime analysis is exploring, inter relating and detecting relationship between the various crimes and characteristics f the crime. This analysis helps in preparing statistics, queries and maps on demand. It also helps to see if a crime in a certain known pattern or a new pattern necessary.

Crimes can be predicted as the criminal are active and operate in their comfort zones. Once successful they try to replicate the crime under similar circumstances. The occurrences of crime depended on several factors such as intelligence of criminals, security of a location,etc The work has followed the steps that used in data analysis, in which the important phases are Data collection ,data classification, pattern identification, prediction and visualization. The proposed framework uses different visualization techniques to show the trends of crimes and various ways that can predicts the crime using machine learning algorithm.

2. CRIME DATA ANALYSIS

Collection and analysis of crime related data are imperative to secure agencies. the use of a coherent methods to classify these data based on the rate and location of occurrences, detection of the hidden pattern among the committed crimes at different times, and prediction of their future relationship are the most important aspects that have to be addressed. One of the most popular approaches is hot spot analysis. Some of the most popular approaches used for this purpose of point pattern analysis and clustering/distances statistics. Another popular approach is the discovery of pattern or trends through various techniques from data mining, text mining and spatial analysis, and self-organizing maps.[1]An crime analysis tool should be able to identify crime patterns quickly and in an efficient manner for future crime pattern detection and action.

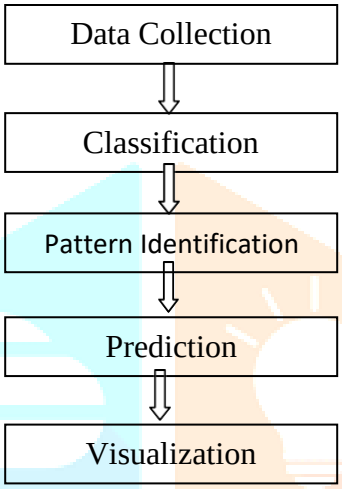
The main purpose of crime analysis is:

- Extraction of crime pattern by crime analysis and based on available criminal information.
- Crime recognition [3].
- Problem of identifying techniques that can efficient and accurate.

s2.1 CRIME ANALYSIS METHODOLOGY

The crime analysis methodologies are:-

- Data Collection
- Classification
- Pattern Identification
- Prediction
- Visualization



Crime analysis steps

Data collection

The data collection is first methodology in crime analysis. Data's are collected from various different websites, news sites and blogs. The collected data is stored into database for further process. This is unstructured data and it is object oriented programming which is easy to use and flexible.

Crime data is an unstructured data since no of field, content, and size of the document can differ from one document to another the better option is to have a schema less database. Also the absence of joins reduces the complexity. Other benefits of using an unstructured database are that:

- Large volume of structured, semi-structured, and unstructured data.
- Object-Oriented programming that is easy to use and flexible.

Classification

In this step use Naive Bayes Algorithm which is supervised learning method. Naive Bayes classifier is a probabilistic classifier which when given an input gives a probability distribution of set of all classes rather than providing a single output. One of the main advantages of the Naïve bayes Classifier is simple, and coverage quicker than logistic regression [2].Compare to other algorithm like SVM (Support Vector

machine) which takes lots of memory.

Using naïve Bays algorithm is create a model by training crime data related to vandalism, murder, robbery, burglary, sex abuse, gang rape, etc. Naive Bayes is that works well for small amount of training to calculate the classification parameter. Estimating probability sometimes while checking a probability P(A) * P(B/D) *P(C/D) * P(E/D) where P(C/D)=0[2].

Pattern Identification

A third step is the pattern identification where we have identify trends and patterns in crime. For finding crime pattern that occurs frequently we are using apriori algorithm. Apriori can be used to determine association rule which highlight general trends in the database. By using pattern identification it will helps to the police officials in an effective manner and avoid the crime occurrences in particular place by providing security, CCTV, fixing alarms etc.

Crime Prediction

The second Approach is predicting the crime type that might occur in a specific location within particular time. To predict an expected crime type is provide four related features of the crime. The features are: occurrence month, the occurrence day of the week, the occurrences time and the crime location. Prediction is stating probability of an event in future period time. A Classification approach is used crime prediction in data miningn[1]classify areas into hotspots and cold spots and to predictive an area will be a hotspot for residential burglary. Variety of classification techniques are used for predicting the crime:-[1]

- K-Nearest Neighbor (k-NN)
- Decision trees (J48)
- Support Vector Machine (SVM)
- Neural Networks
- Naïve Bayes and ensemble learning

Linear Regression methods are also used for predicting the crime prediction. Based on the crime probability. The formula for a regression line is

Y=aX + b where, Y is the predicted score, b is the slope of the line, and A is the Y intercept. b = r sx\sy

And the intercept (A) can be calculated as A=MY –bMX.

Some Theories are used to predicting the crimes are:

- Integrated theory
- Biological theory
- Psychological theory
- Sociological theory
- Conflict theory
- Victimization theory
- Choice theory

Visualization

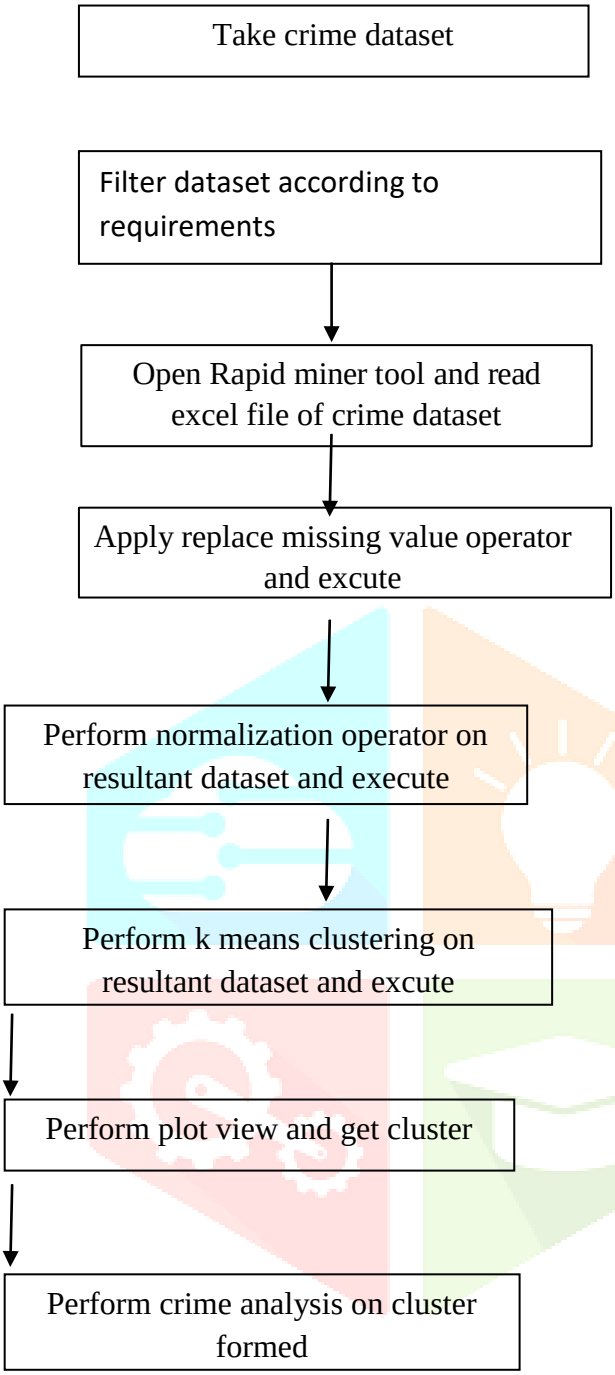
The crime prone area can be graphically reoresented using a heat amp which indicates level of activity,dark colour indicates low activity and brighter colour indicates the high activity.

Advantages of using heat map are [2].

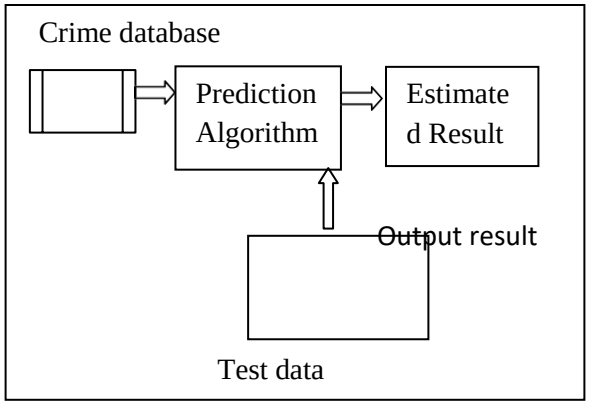
- Numerical and category based color images
- Gradient color range
- Analyze only the data we want
  - Out of range data is automatically discarded.

Map showing crime prone area

FLOW CHART OF CRIME ANALYSIS AND PREDICTION



Flow chart of crime analysi



Predicting future crime trends [3]

3. DATA

This dataset contains a record of incidents that the Austin Police Department responded to and wrote a report. Data is from 2003 to present. This dataset is updated weekly. Understanding the following conditions will allow you to get the most out of the data provided. Due to the methodological differences in data collection, different data sources may produce different results. This database is updated weekly, and a similar or same search done on different dates can produce different results.

Comparisons should not be made between numbers generated with this database to any other official police reports. Data provided represents only calls for police service where a report was written. Totals in the database may vary considerably from official totals following investigation and final categorization. Therefore, the data should not be used for comparisons with Uniform Crime Report statistics. The Austin Police Department does not assume any liability for any decision made or action taken or not taken by the recipient in reliance upon any information or data provided. Pursuant to section 552.301 (c) of the Government Code, the City of Austin has designated certain addresses to receive requests for public information sent by electronic mail.

Summary of the crime Report dataset Rows: 2.21M

Columns:	27

In the dataset contains different types of crimes (attributes) are considered like murder, rape, kidnapping, dacoit, robbery, burglary, cheating, dowry deaths, arson etc.

No	Crime type	No of records
1	Anti social behavior	44,070
2	Burglary	22,081
3	Violent crime	21,333
4	Other theft	19,538
5	Vehicle crime	18,260
6	Drugs	8,336
7	Roberry	2,166
8	Theft from person	799

Types of crimes [1]

4. ALGORITHMS

Our experiment choose the algorithm are

- Instance based algorithm
- Decision tree
- Linear regression
- K-means algorithm

1. Instance Based Algorithm

-The instance based algorithm is also called as tge machine based learning is a family of learning algorithm that, instead of performing explicit generalization, compares new problems instances with instance seen in training, which have been stored in memory. These stored their training set when predicting a value or class for a new instances, they compute distance training instances to make a decision.

The algorithm in this category for numerical prediction can divided into two types: similarity- based, e.g., Euclidean or entropy based and regression-based e.g., LWL Since regression is one of the most popular methods for numerical prediction[1].

The advantages of the Instances based Algorithm is it over other methods of machine learning is its ability to adapt its model of machine learning is its ability to adapt its model to previously unseen data. Instance based learners may simply store a new instance or throw an old instance away. The Disadvantages of the instances based Algorithm are its need more storage and computational complexity.

2. Linear Regression

-It is simple form of regression. Linear regression attempts to model the relationship between the two variables by fitting a linear equation to observe the data. this is widely used in statistics. For this purpose ,linear functions are used for which the unknown parameter i.e., weight of the independent variables, are estimated from the training data[1].this can be used to predict the values One of the most common estimating method is least mean square.

Linear regression algorithms for predicting include simple regression multiple regression and pace regression, which is suitable for data of high dimensionality and only accepts binary nominal attributes.[1].

The main advantages of the linear regressions is gain a far greater understanding of the variables that can impact its success in the coming weeks,months and years into the future.The disadvantages of the regression

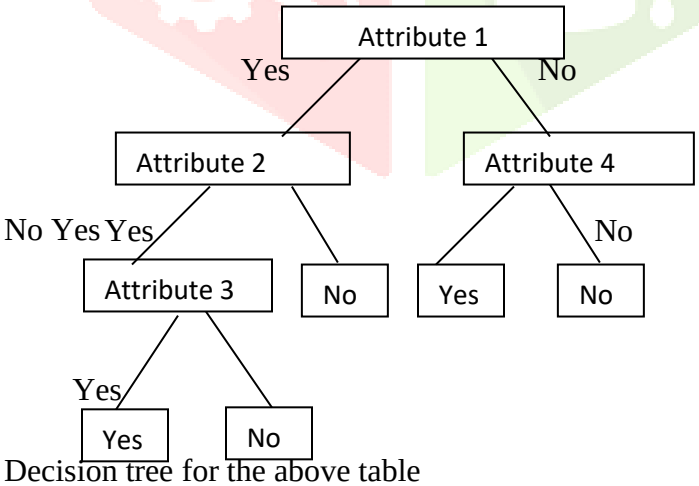
is its linearity.If the data has non linear dependencies, a linear regression model will ouput the best fitting line which may not fit very well.

3. Decision Tree

Decision tree is used for both the prediction and classification. for the classification purpose a function can be learned this is intervals defined by splits on the individuals attributes value

Area Sensitivity	**Notable event**	**VIP Presence**	**Criminal group**	**Crime**
Yes	Yes	Yes	No	Yes
Yes	Yes	No	Yes	No
No	No	No	Yes	No
Yes	No	No	No	No
Yes	Yes	Yes	Yes	Yes
No	Yes	No	No	No

TABLE: Decision Tree for Delhi [2]



Decision tree for the above table

A Root node, that has incoming edges and zero or more outgoing edges.

- Internal nodes, each of which has one incoming edges and two or more outgoing edges.
- Leaf node or end node, each of which has exactly one incoming edge and no outgoing edges.

For prediction purpose, the decision trees algorithm for classification have been adapted to output a numerical value the main difference

Is that the leaves of the tree have numerical values, unlike classification trees have class labels.

Advantages of the decision trees are It is very simple to understand and help determine worst,best and expected values for different scenarios.it can be combined with other decision techniques.Some of the Disadvantages of the Decision tree are They are unstable, They are often relatively inaccurate, Calculation can get vey complex.

4. K-Means Algorithm

:K –means is the simplest and most commonly used portioning algorithm among the clustering algorithm in scientific and industrial software[3].Acceptance of k means is mainly due to its being simple .This algorithm is also suitable for clustering of a large datasets since it has much less computational complexity grows linearly by increasing of the data points.

Advantages of the k-means algorithm are relatively simple to implement, Scales to large dataset, Guarantees convergence, easily adapts to new examples. Disadvantages of the k-means algorithm are Choosing manually, Being dependent on initial values, clustering data of varying sizes and density.

LITERATURE SURVEY

SNO	TITLE	AUTHOR	JOURNAL	OVERVIEW OF PAPER	METHODS USED	ABSTRACT THEAM	ACCURACY RESULT OBTAINED	FUTURE WORK
1	An Exploration of Crime Prediction Using Data Mining on Open Data	Ginger Saltos and Minhaela Cocea	International Journal of Information and Technology& Decision Making (World Scientific)	Crime data has been systematically recorded by the police for many years and in the last decades, there has been a surge of Open Crime Data and of apps or web based application displaying crime statistics on maps, both by official sources, such as from police UK, and other sources usingthe same official data. In this paper investigate predict many types of crime and discuss their applicability.	Crime prediction, Data mining, Open data, regression, decision tress, instance-based learning.	This Paper explore models predicting the frequency of several types of crimes. in this paper using three types of algorithms they are instance-based learning, Decision trees, regression.	The experiments were conducted using the SCIAMA High Performanc e Computer Cluster at the University of Portsmouth and the Weka software	Further experiment can be conducted to investigate other aspects such as time frame for prediction, the amount of data necessary for reliable prediction models, and predictive models for particular types of crime.
2	Crime Analysis and Prediction Using Data Mining	Shiju Sathyadevan,Deva n M.S, Surya Gangadharan	First International Conference on networks & soft computing	According to the Crime record Bureau crimes like burglary, arson, etc have been decreased while crimes like murder, sex, abuse, gang rap	Naïve bayes, Apriori algorithm, Decision tree, NER, Mongo Db, GraphDBs	In this paper predict the crime based on the occurrence s of the crimes and	This paper has tested the accuracy of classification and prediction based on different test	Our system predicts crime prone regions in India on a particular day. It will be more

				etc have been increased. In this paper data's are collected from various sources like websites, blogs, news sites, social media, RSS feeds etc. This huge data is used as a record for creating a crime record database. The crime analysis using five steps they are Data Collection, Classification, Pattern identification, prediction, and Visualization.		can visualize crime prone areas. Using the concept of data mining can extract previously unknown, useful information from an unstructure d data.	sets. Classificatio n is done based on the Bayes theorem which showed more than 90% accuracy.	accurate if we consider a particular stage/regio n. Also another problem is that we are not predicting the time in which the crime is happening .
3	Crime Detect ion Techni ques Using data Minin g and K-Means	Khushab u A.Bokde, Tisksha P. Kakade, Dnyanes hwari S. Tumsare, Chetan G. Wadhai B.E Student	Internatio nal Journal of Engineeri ng Research & Technolog ieEngineer ing (IJETER)	This paper is mainly focusing on crime Analysis, Clustering and Clustering by K-means algorithms methods. some of the purpose of crime analysis are Extraction of crime patterns by crime analysis and based on available criminal information, crime recognition. Clustering means	Crime, Clustering , K-Means Algorithm	The main objective of this paper is to classify clustered crimes based on occurrence frequently during different years. Data mining is used to extensively in. terms of analysis,	From the accuracy result, crime data mining has a promising future for increasing the effectivenes s and efficiency of criminal and intelligence analysis .	From the clustered result it is easy to identify crime trends over year and can be used to design precaution methods for future.

				Division of a set of data or objects to a number of clusters. There by cluster is composed of a set of similar data which have same as a group. K- means is the simplest and most commonly sued portioning algorithm among the clustering algorithms in scientific and industrial software.		investigaga tion and discovery of patterns for occurrence s of crime.		
4	Survey on crime analysi s and predict ion using data mining techni ques.	Benjamin Fredrick David. H and Suruliand i	ICTACT Journal on Soft computing	Criminology is process that is used for identify the crime characteristics and crime identify. By using this data mining algorithms will be able to produce crime reports and help in the identification of criminal much faster than any human could. The criminals when leaving the crime scene does leave some trace which can be used as	Criminolo gy, Crime Analysis, Crime Prediction , Data Mining	The extraction of the new information is predicted using the existing datasets. The criminals can also be predicted based on the crime data. The main aim of this work is to perform a survey on the	The quantitative analysis produced results which shows the increase in the Accuracy level of classificatio n because of using the GA to optimize the parameters.	In future ,there is a plan for applying other classificati on algorithm on the crime data and improving the accuracy in prediction.

				clue to identify the criminals .This process is used to identifying the criminals based on the clues or information given by the local peoples. in this paper using some criminal analysis methods they are Text, content and NPL-based methods, Crime patterns .		supervised learning and unsupervis ed learning techniques that has been applied towards criminal identificati on.		
5	Crime Pattern Analys is, Visual ization s And Predict ion Using Data Minin g	Tushar Sonawan ev, Shirin Shaikh,S haista Shaikh, Rahul Shinde, Asif Sayyad.	Indian Journal of Computer Science and Engineeri ng (IJCSE)	Crime analysis is exploring, inter relating and detecting relationship between various crimes and characteristics of crimes. The main aim of this research paper consists of developing analytical data mining methods that can systematically address the complex problems related to various form of crime. From this literacy study, it	K-Means, Cluster, Correlatio n	Preventing are taken to reduce the increasing number of cases of crime against women. A huge amount of data set is generated every year on the basis of reporting the crime.	Result will be in the form of correlation between various crime and location of crime. crime can also be correlated on the basis of age ,group, location of crime & type of crime.	The future scope is prediction of the crime will be displayed using various diagram pie charts, heat maps, spikes and graphs.

				could be concluded that crime details increasingly to very large quantities running into zota bytes.				
6	Crime Analysis And prediction using Data Mining Techniques	Rajkumar.S, Sakkarai Pandi.M, Soundarya Jagan.J,Varnikasree.P	International Journal of recent trends in engineering and reaserch	Main contribution of this paper is to propose a new approach based on deep learning success in different classification task such object detection image recognition natural image processing and dimensionality reduction deep learning algorithms use deep architectures of multiple layers to extract features from raw data.	Data mining, machine learning, crime analysis, crime prediction.	Crime hotspot prediction has previously been suggested. Crime hotspot prediction leverages past data in order to identify crime hotspots, or social media data.	Analyzed and compared different algorithm on crime data determine which algorithm performs better of crime prediction.	As a future extension of our work, we plan to apply more classification models to increase crime prediction accuracy and to enhance the overall performance.
7	Systematic Review of Crime Data Mining	Sapreet kaur,Dr. Williamjeet Singh	International Journals of Advanced Research in Computer Science (IJARCS)	This paper is explains techniques used, challenges addressed, methodologies used, and crime data mining and analysis paper. The methodologies is	Crime data mining, crime data analysis, systematic review, systematic study	Crime data mining has the ability of extracting useful information and hidden patterns from the large data	The results of this result may help new potential users in understanding the range of available crime data mining	In future to improve the performance of these classification. Hence, the usage of other classification

				composed of three stages the first stage involves the research work related to crime data mining, second stage is concerned with establishing a classification and the third stage is involves the presentation of summary of research in crime data mining and analysis and report of this survey		sets. The crime data mining challenges are becoming stimulating opportuniti es for the coming years.	technologies and technologies .	algorithm like machine learning will be explored in future.
8	Survey paper on Crime Predict ion using Ensem ble Appro ach	Ayishesh im Almaw,K alyani kadam	Internatio nal Journal of Pure and Applied Mathemati c	Different classification algorithm in problem resolving are nominated based on the suitable requirements in crime data prediction.One technique may provide better accuracy value than different techniques which are nominated for solving the particular problems .Some papers introduced combined different models to achieve better	Crime prediction, naïve Bayes, J48,Artific ial network	This paper investigates a number of data mining algorithms and ensemble learning which are applied on crime data mining. This paper delivers reasonable investigatio n of data mining techniques	A Survey is conducted so that crime forecasting can be improved by the use of efficient data collection and data mining strategies.	In future predicts what kind of crime might occur next in a particular district within a certain periods of time and identifier the season and time factor at which crimes are occur more

				performance which overcome the individual models to achieve better performance which overcomes the individual models called ensemble learning.		and ensemble classificati on techniques for discovery and prediction of upcoming crime.		frequently happening crime.
9	Revie w on Crime Analys is and Predict ion Using Data Minin g Techni ques	Dr .M Sreedevi, A.harsha vardhan reddy, Ch.venka ta Sai Krishna Reddy	Internatio nal Journal of Innovation Researchi ng Science, Engineeri ng and Technolog y	Clustering techniques is used to existing and known crimes. It is able to detect newer and unknown patterns in future .this paper collect the dataset from crime data available in national crime bureau of records. by using the correlated data set we can identify the correlated crimes. And the k means algorithm is used for group the data into same characteristics.	Data mining, machine learning, crime analysis, crime prediction	Extract previously unknown useful information from an unstructure d data. This paper explains various types of criminal analysis and crime prediction using several data mining techniques	By observing this survey this is a theoretical study for several methods and methodologi es in identificatio n of crime and criminals which includes text/NLP, crime patterns, geo location, prisoner methods, communicat ion based methods data collection,	In future more data mining techniques are used for predicts the crime pattern.

							classificatio n, pattern identificatio n, prediction.	
10	Crime analysi s Journa l of engine ering & techno logy	Deepika k,k, Smitha Vinod	Internatio nal journal of Engineeri ng & Technolog y	According to this paper burglary and robbery has reduced over a period of 53 tears by79.84%and 28.85% respectively. crime like murder and kidnapping has hiked by 7.39% and 47.80% respectively. Crime analysis is a part of criminology plays an important role in crime detection .data mining helps in solving the crimes faster and this techniques give good results when applied on crime dataset, the information obtained from the data mining techniques can help the police department.	Clustering , classificati on, Visualizati on, k-means, random forest, Neural networks	The approach consists of the following steps – data preprocessi ng, clustering, classificati on, and visualizatio n. Data mining techniques are often applied to criminolog y as it provides good results.	A good accuracy of 99.93% is obtained and this verifies the correctness of the instances.	The future enhanceme nt of this research work focuses on training bots to predict the crime prone area by using machine learning techniques.

CONCLUSION

In this paper focused on building predictive models for crime frequencies per crime type per month. The crime rates in India are increasing day by day due to many factors such as increase in poverty, implementation, corruption, etc. The proposed model is very useful for both the investigating agencies and the police official in taking necessary steps to reduce crime. The project helps the crime analysis to analysis these crime networks by means of various interactive visualization.

Future enhancement of this research work on training bots to predict the crime prone areas by using machine learning techniques. Since, machine learning is similar to data mining advanced concept of machine learning can be used for better prediction. The data privacy, reliability, accuracy can be improved for enhanced prediction.

REFERENCE

[1] Ginger Saltos and Mihaela Coacea, An Exploration of Crime prediction Using Data Mining on Open Data, International journal of Information technology & Decision Making,2017.

[2] Shiju Sathyadevan, Devan M.S, Surya Gangadharan.S, Crime Analysis and Prediction Using Data Mining, First International Conference on networks & soft computing (IEEE) 2014.

[3] Khushabu A.Bokde, Tisksha P.Kakade, Dnyaneshwari S. Tumasare, Chetan G.Wadhai B.E Student, Crime Detection Techniques Using Data Mining and K-Means, International Journal of Engineering Research & technology (IJERT) ,2018

[4] H.Benjamin Fredrick David and A.Suruliandi,Survey on crime analysis and prediction using data mining techniques, ICTACT Journal on Soft computing, 2017.

[5] Tushar Sonawanev, Shirin Shaikh, rahul Shinde, Asif Sayyad, Crime Pattern Analysis, Visualization And prediction Using Data Mining, Indian Journal of Computer Science and Engineering (IJCSE), 2015.

[6] RajKumar.S, Sakkarai Pandi.M, Crime Analysis and prediction using data mining techniques, International Journal of recent trends in engineering & research,2019.

[7] Sarpreet kaur, Dr. Williamjeet Singh, Systematic review of crime data mining, International Journal of Advanced Research in computer science , 2015.

[8] Ayisheshim Almaw, Kalyani Kadam, Survey Paper on Crime Prediction using Ensemble Approach, International journal of Pure and Applied Mathematics,2018.

[9] Dr .M.Sreedevi, A.Harha Vardhan Reddy, ch.Venkata Sai Krishna Reddy, Review on crime Analysis and prediction Using Data Mining Techniques, International Journal of Innovative Research in Science Engineering and technology ,2018.

[10] K.S.N .Murthy, A.V.S.Pavan kumar, Gangu Dharmaraju, international journal of engineering, Science and mathematics, 2017.

[11] Deepiika k.K, Smitha Vinod, Crime analysis in india using data minig techniques , International journal of Enginnering and technology, 2018.

[12] Hitesh Kumar Reddy ToppyiReddy, Bhavana Saini, Ginika mahajan, Crime Prediction &Monitoring Framework Based on Spatial Analysis, International Conference on Computational Intelligence Data Science (ICCIDS 2018).