Crime Prediction And Analysis Using Data Mining

[1]Abhedya Bhole, [2]Ankit Chaudhari, [3]Sanket Nagla, [4]Priyanka Jadhav
Prof. Kanchan Dhomse

Department Of Computer EnfineeringMET Bhujbal Knowledge City Institute Of Engineering,Adgaon Nashik

----

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. 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

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 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 theytry 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.

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:

1. Extraction of crime pattern by crime analysis and based on available criminal information
2. Crime recognition
3. Problem of identifying techniques

CRIME ANALYSIS METHODOLOGY

The crime analysis methodologies are:-

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

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 isunstructured 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. 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

Linear Regression methods are also used for predicting the crime prediction Based on the crime probability.The formula for a regression is predicted score$Y=aX + b$ where, Y is the line, b is the slope of the line

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,darkcolour indicates low activity and brighter colour indicates the high activity.

Flow chart of crime analysis
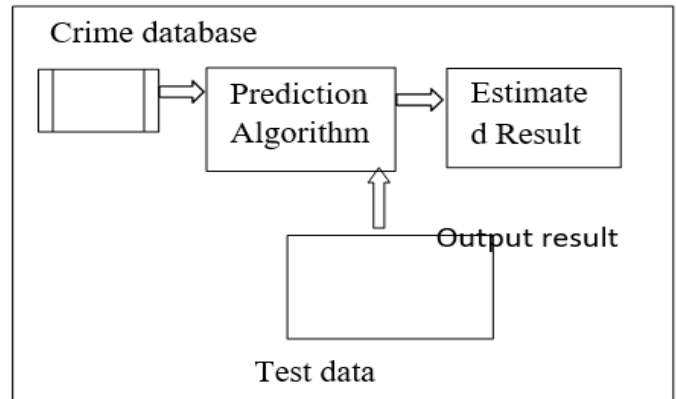
1. DATA

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.

ALGORITHMS

Our experiment choose the algorithm are

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

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 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.

1. 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 estimatingmethod 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 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 fittingline which may not fit very well.

1. Decision Tree:

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 very complex.

K-Means Algorithm:

K –means is the simplest and most commonly used portioning algorithm among the clustering algorithm in scientific and industrial software 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.

CONCLUSION:

In this paper focused on building predictive models for crime frequencies per crime type per month. Thecrime rates in India are increasing day by day due to many factors such as increase in poverty, implementation, corruption, etc. The proposed model 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 thepolice 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.

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

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

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

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