Algorithms to Predict Crime Pattern to Prevent Future Crimes

Shital B. Jadhav[1], Neelam Rajpal[2], Heena Mulla[3], Nidhi Patel[4], Pooja[5]

Abstract- *Crime around the world has become problem of every nation, around the globe many countries are trying to curb this problem. Crime analysis is a law enforcement function that involves systematic analysis for identifying and analyzing patterns and trends in crime and disorder. Crime analysis also plays a role in devising solutions to crime problems, and formulating crime prevention strategies. A huge amount of data set is generated every year on the basis of reporting of crime. This data can prove very useful in analyzing and predicting crime and help us prevent the crime to some extent. Crime analysis is an area of vital importance in police department. Dataset is classified on the basis of some predefined condition. Using data mining technique data can be predicted and visualized in various form in order to provide better understanding of crime patterns.*

I. INTRODUCTION

India is a vast country with diversified societies. Criminology is scientific study of crime and criminal behaviour in order to detect crime characteristics. Use of data mining techniques can produce important results from crime dataset.The very step in study of crime is crime analysis .Crime analysis is exploring, inter relating and detecting relationships between various crimes and characteristics of crimes .Police department maintains crime data at the record. This data contains huge amount of data set with complex relationships which needs use of data mining techniques in order to be transformed into useful information. The knowledge extracted from the dataset can be a great tool & support to the police department to prevent crimes.

The Cambridge Police Department has one of the oldest crime analysis units in the world and their historical data was used to train Series Finder to detect housebreak patterns. The algorithm tries to construct a modus operandi (MO). The M.O. is a set of habits of a criminal and is a type of behavior used to characterize a pattern. The data of the burglaries include means of entry (front door, window, etc.), day of the week, characteristics of the property (apartment, house), and geographic proximity to other break-ins. Using nine known crime series of burglaries, Series Finder recovered

most of the crimes within these patterns and also identified nine additional crimes. Machine learning is a tremendous tool for predictive policing. If patterns are identified the police can immediately try to stop them. Without such tools it can take weeks and even years of shifting though databases to discover a pattern. Series Finder provides an important data-driven approach to a very difficult problem in predictive policing.

Data mining techniques have higher influence in the fields such as, Law and Enforcement for crime problems, crime data analysis, criminal career analysis, bank frauds and other critical problems. In recent years, data clustering techniques have faced several new challenges including simultaneous feature subset selection, large scale data clustering and semi-supervised clustering. Mostly, cluster analysis is an important human activity which indulge from childhood when learn to distinguish between animals and plants, etc by continuously improving subconscious clustering schemes. It is widely used in numerous applications including pattern recognition, data analysis, image processing, and market research etc.

Before this clustering algorithms have been used for crime analysis. For example, one site it is revealed that suspect is literate and from next site/witness it is revealed that suspect is youth etc. By describing the offender details it gives a complete picture from different crime incidents. Today mostof it is manually done with the help of multiple reports that the detectives usually get from the computer data analysts and their own crime logs.The reason for choosing this method is that we have only data about the known crimes we will get the crime pattern for a particular location. The nature ofcrimes change over time, so we are using clustering technique in order to detect newerand unknown patterns.

II. LITERATURE REVIEW

J. Agarwal, R. Nagpal and R. Sehgal have analyzed crime and considered homicide crime taking into account the corresponding year and that the trend is descending from 1990 to 2011. They have used the k-means clustering technique for extracting useful information from the crime dataset using RapidMiner tool.

k-means clustering is a method that is popular for cluster analysis in data mining.

k-means clustering aims to partition n observations into k clusters in which each observation belongs to the cluster with the nearest mean, serving as a prototype of the cluster. This results in a partitioning of the data space into Voronoi cells.

Priyanka Gera and Dr. Rajan Vohra in have used a linear regression for prediction the occurrence of crimes in Delhi (India). They review a dataset of the last 59 years to predict occurrence of some crimes including murder, burglary, robbery and etc.

Linear regression is an approach for modeling the relationship between a scalar dependent variable and one or more independent variables.

Brown (1998) constructed a software framework called ReCAP (Regional Crime Analysis Program) for mining data in order to catch professional criminals using data mining and data fusion techniques. Data fusion was used to manage, fuse and interprets information from multiple sources. The main purpose was to overcome confusion from conflicting reports and cluttered or noisy backgrounds. Data mining was used to automatically discover patterns and relationships in large databases.

De Bruin et. al. (2006) introduced a framework for crime trends using a new distance measure for comparing all individuals based on their profiles and then clustering them accordingly.

Shiju Sathyadevan and Surya Gangadharan. S used SVM and Naïve bayes for classification of crimes. Naïve bayes used is more efficient than SVM as it has high performance, it is easy to implement and gives more accuracy even at less training dataset. It also handles zero frequency problems which SVM cannot.

For our project we have choosen naïve bayes and K-means instead of SVM because of their simplicity and good performance. SVM i.e support vector method takes lots of memory than Naive bayes algorithm.

Advantages of K-means: K-means clustering algorithm is easy to implement and apply even on large data sets.If the variables are huge, then k-means is faster than hierarchical clustering, if we keep k smalls.Unlike hierarchical clustering, k-means clustering produce tighter clusters especially if the clusters are globular.

Advantages of Naive Bayes: Super simple, you're just doing a bunch of counts. If the NB conditional independence assumption actually holds, a Naive Bayes classifier will converge quicker than discriminative models like logistic regression, so you need less training data. And even if the NB assumption doesn't hold, a NB classifier still often does a great job in practice.

III. EXISTING SYSTEM

At present there is only a stand alone system present for analysis of crime pattern and prediction of crime prone areas.For the accessing of any information, there is a system used indian police departmnets which can share the files of crime among police department.The Common Integrated police Application(CIPA) is an application to automate the process(workflow)at primary sources of data itself .It was designed and developed by NIC with multilingual Interface. But these system cannot predict the crime prone areas.

- Person has to manually analyse the piles of files which results in more time consumption and needs storage of hard copies.secondly you cannot access it everywhere.
- Maintenance of such a large amount of data and searching and making prediction from that data is not feasible manually.
- There has been an enormous increase in the crime in the recent past.to solve this social issue,many different systems are developed but they cannot predict the crime prone areas

The existing stand alone system is a peer to peer application.

- It has no central server.
- Each workstation on the network shares its files equally with the others. There's no central storage or authentication of users.it may leads to major security issue.
- They applications are more difficult to update and the storage capacity is restricted to the capacity of storage device.
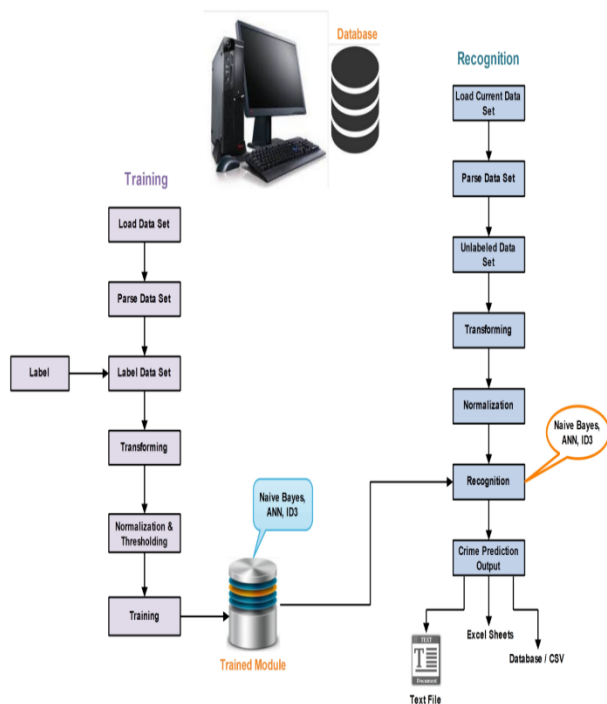
Existing architecture

There are steps in doing Crime Analysis:

1) Data Collection

In data collection step data from different web sites like news sites, blogs, social media etc. are collected and is stored into database for further process.The database used is mongodb as it is unstructured.

2) Classification

For classification Naïve Bayes a supervised learning method is used. Naive Bayes classifier gives a probability distribution of set of all attributes instead of providing a single output.

3) Pattern Identification

In pattern identification phase trends and patterns in crime are identified. For this phase apriori algorithm is used.it is used to determine association rule that highlight general trends in the database.

4) Prediction

For prediction decision tree concept is used. An internal node of decision tree represents test on an attribute, and each branch represents outcome of a test.

5) Visualization

The crime prone areas is represented using a heat map or bar graphs.

IV. PROPOSED SYSTEM

- In the proposed system, we are introducing the application which will predict the crime that criminals can do in the future and the crime prone areas.
- As our system is client based it provides us these different features which the existing standalone system don't.

1) Centralization :Servers help in administering the whole set-up. Access rights and resource allocation is done by Servers.

2) Proper Management : All the files are stored at the same place. In this way, management of files becomes easy. Also it becomes easier to find files.

3) Back-up and Recovery possible: As all the data is stored on server its easy to make a back-up of it. Also, in case of some break-down if data is lost, it can be recovered easily and efficiently. While in peer computing we have to take back-up at every workstation.

4) Upgradation and Scalability :
Changes can be made easily by just upgrading the server. Also new resources and systems can be added by making necessary changes in server.

5) Accessibility : From various platforms in the network, server can be accessed remotely.

6) Storage Efficiency: As new information is uploaded in database , each workstation need not have its own storage capacities increased (as may be the case in peer-to-peer systems). All the changes are made only in central computer on which server database exists.

7) Security : Rules defining security and access rights can be defined at the time of set-up of server.

Servers can play different roles for different clients.

- We will perform mining on all the previously stored criminal records and will calculate the possibility and predict the next type of crime and next area that may get affected by crime.
- This prediction will be done on the basis of attributes like criminal record, education, occupation,family background and other various attributes.
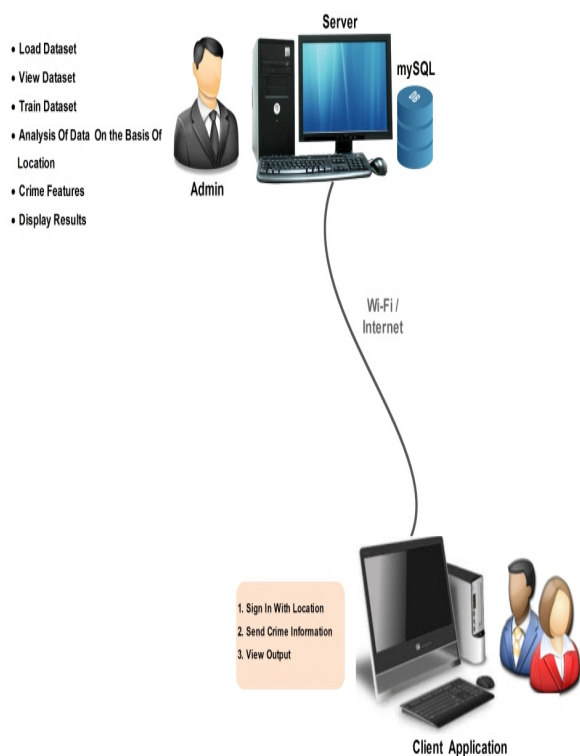
V. METHODOLOGY

1) Data Collection

In data collection step data from different web sites like news sites, blogs, social media etc. are collected and is stored into database for further process.The database used is mysql as it is open source relational database.we will process the data and will find the frequently occuring crimes.then we will relate this crime type to location.

2) Transformation:

In this phase we will perform clustering. The reason for choosing this method is that if we have data of known crimes we can get a particular place's crime pattern. But the criminal data in the database keeps on changing so inorder to handle new and unknown patterns clustering is a good technique to work with.

3) Training:

In this phase we will apply naïve bayes algorithm on the transformed data set we get from transformation step.we have choosed Naive Bayes because it is the basis for many machine-learning and data mining methods. The algorithm is used to create models with predictive capabilities. It provides new ways of exploring and understanding data.

1) it can handle high amount of data with ease.
2) It gives more efficient output as compared to other methods.

4) Prediction

In the prediction step the transformed current value and training data set will be analysed using frequent count method. And at the end of this phase frequently occurring crime type will be predicted and stored into database.

5) Pattern Recognisation

In pattern recognisation phase we will again apply naïve bays algorithm to training dataset to find the frequently occuring value of location attribute and crime type attribute on the basis of output of prediction phase and pattern recognisation phase we will get the final output.

6) Visualization

The crime prone areas will be then represented using maps or bar graphs.

VI. CONCLUSION AND FUTURE WORKS

The present scope of our project is prediction of crime type that what type of crime may happen in future and crime prone areas i.e. next location where the crime may take place. we can also predict the estimated time for which the crime to take place as a future scope Of our project. Along with this, one can try to predict identity of an individual criminal who may perform the crime. we will do clustering on different data sets. This will help system to learn change in pattern of crime and location automatically. Our software will predict the crime prone area i.e. location where chances of happening of next crime is maximum. we will use naïve bays algorithm for this purpose. this will find the crime prone areas. Instead of this web based application, an android application can be developed for this as a future so that it can easily be accessible anywhere.

REFERENCES

[1] Malathi. A and Dr. S. Santhosh Baboo.,"an enhanced algorithm to predict a future crime using data mining. "International Journal of Computer Applications,6, May 2011. Published by Foundation of Computer Science.

[2] Tong Wang, Cynthia Rudin, Daniel Wagner, and Rich Sevieri. "Detecting patterns of crime with series finder. "In Proceedings of the European Conference on Machine

Learning and Principles and Practice of Knowledge Discovery in Databases ,2013.

[3] Clifton Phua,Kate Smith-Miles,Vincent Cheng-Siong Lee, and Ross Gayler. "Resilient Identity Crime Detection",IEEE transactions on knowledge and data engineering, vol. 24, no. 3, March 2012.

[4] Shiju Sathyadevan, Devan M.S,Surya gangadharan .S."crime analysis and prediction using data mining",IEEE 2014 international conference on network and soft computing.

[5] Rasoul Kiani ,Siamak Mahdavi and Amin Keshavarzi "Analysis and Prediction of Crimes by Clustering and Classification" (IJARAI) International Journal of Advanced Research in Artificial Intelligence, Vol. 4, No.8, 2015

[6] Mrs. M.Kavitha [1], Ms.S.T.Tamil Selv."Comparative Study on Apriori Algorithm and Fp Growth Algorithm with Pros and Cons "International Journal of Computer Science Trends and Technology (I JCS T) – Volume 4 Issue 4, Jul - Aug 2016

[7] Mrs. Prof. Jadhav Shital ,Ms. Kardile Bhagyashri,Ms .Borikar Utkarsha Sudhir, Ms. Vasekar Vrushali Vasant,Ms. Aphale Madhuri Sudam"Forensic Investigation for Database Tamperingusing Audit Logs"International Journal of Engineering Research & Technology (IJERT)Vol. 4 Issue 03, March-2015.