 ISSN: 2321-8169
124 – 127

Crime Rate Prediction using KNN

Ms. Vrushali Pednekar
Computer Department
Saraswati CollegeOf Engineering
Kharghar,Navi Mumbai
vrush.p96@gmail.com

Ms. Trupti Mahale
Computer Department
Saraswati College Of Engineering
Kharghar,Navi Mumbai
trupti.mahalenimba@gmail.com

Ms. Pratiksha Gadhave
Computer Department
Saraswati College Of Engineering
Kharghar,Navi Mumbai
pratiksha.gadhave@gmail.com

Prof. Arti Gore
Computer Department
Saraswati College Of Engineering
Kharghar,Navi Mumbai
abs_14284@yahoo.co.in

***Abstract*—**Crime is one of the most predominant and alarming aspects in our society and its prevention is a vital task. Crime analysis is a systematic way of detecting and investigating patterns and trends in crime. Thus, it becomes necessary to study various reasons, factors and relationship between different crimes that are occurring and finding the most appropriate methods to control and avoid more crimes. The main objective of this project is to classify clustered crimes based on occurrence frequency during different years. Data mining is used broadly in terms of analysis, investigation and discovery of patterns for occurrence of different crimes. In this work, various clustering approaches of data mining are used to analyze the crime data. The K-Nearest Neighbour (KNN) classification is used for crime prediction. The proposed system can predict regions which have high probability for crime rate and can forecast crime prone areas. Instead of focusing on causes of crime occurrence like criminal background of offender, political enmity etcit will focuse mainly on crime factors of each day.

Keywords*: Crime, clustering, classification, genetic algorithm, k-nearest neighbor (KNN).*

*****

I. INTRODUCTION

Day by day the crime rate is increasing considerably. Crime cannot be predicted since it is neither systematic nor random. Also the modem technologies and hi-tech methods help criminals in achieving their misdeeds. According to Crime RecordsBureaucrimeslikeburglary,arsonetchavebeendecreased whilecrimeslikemurder, sex abuse, gang rape,etc have been increased. Even if system cannot predict the victims of crime but can predict the place that has probability for its occurrence.

Today, collection and analysis of crime-related data are imperative to security agencies. The use of analytical method to classify these data based on the rate and location of occurrence, detection of the hidden pattern among the occured crimes at different times, and prediction of their future relationship are the most important aspects that have to be addressed. So we need methodologies to predict and prevent crime.DataMiningprovidesclusteringandclassificationtechnique forthispurpose. Clustering is used for grouping the similar patterns to identify crimes. Cluster refers toageographicalcollectionofcrimethatcanbevisualizedusingtheg eo-spatialplot in the map. Clustering in crime is mainly used to identify the patterns in crime and also used to predict the crime. Classification is a technique of data analysis that is used to extract and predicts future trends in data based on similarity measures.

There are steps in doing Crime Analysis:

1) Data Collection
2) Classification
3) Pattern Identification
4) Prediction
5) Visualization

1) Data Collection:

Large amount of crime data is collected at police records. This data is made available by National Crime Bureau of Records. This data is in the form of number of cases recorded all over the nation throughout the year. The data is in unprocessed form and contains some wrong as well as missing values. Hence preprocessing of data is crucial task in order to bring the data in proper and clean form. Pre-processing of data includes data cleansing and Preprocessing.

2) Classification:

Thedataset is classified into various groups based on certain characteristics of the data object.Grouping of crimes is done according to states & cities. Classification of the crime is done on the basis of different types of crime. K-mean algorithm can be used to group or cluster data with similar characteristics.

3)Pattern Identification

In these phaseproposed system have to identify trends and patterns in crime. The result of this phase is the crime pattern for a particular place. Here corresponding to each location we take the attributes of that place like weather attributes, area sensitivity, notable event, presence of criminal groups etc. Information regarding patterns helps police officials to facilitate resources in an effective manner.

4) Prediction

Corresponding to each place it builds a model. So for getting the crime prone areas we pass current date and current attributes into the prediction software. The result is shown using some visualization mechanisms.

5)Visualization

The crime prone areas can be graphically represented using a heat map which indicates level of activity, usually darker colorsto indicate low activity and brighter colors to indicate high activity.

II. LITERATURESURVEY

The proposed system introduces a new framework amongstthe previous system for classification and prediction of crime based on real dataset. The purpose of new framework are :

- Generation of training and testing data
- Removing low value attribute using weighting technique to deal with high dimensional data challenge.

A lot of work has been done for predicting crimes which are as follows:

1. S. Sathyadevan, and S. Gangadharanin[6]researcheshaveintroducedcrimeanalysisand prediction usingdatamining. They have proposed an approach between computer science and criminal justice to develop a data mining procedure that can help solve crimes faster. Also they have focused on causes of crime occurrence like criminal background of offender, political, enmity and crime factors of each day.

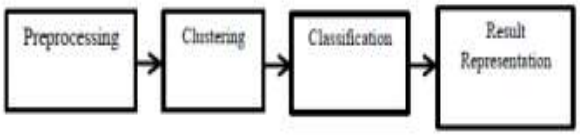

Figure 1. System design.[6]

2. An improved method of classification algorithms for crime prediction has proposed by A. Babakura, N. Sulaiman and M. Yusuf in [4]. They have compared Nave Bayesian and Back Propagation (BP) classification algorithms for predicting crime category for distinctive state in USA. In the first step phase,themodelisbuiltonthetrainingandinthesecondphasethemodelisapplied. The performance measurements such as Accuracy, Precision and Recall are used for comparing of the classification algorithms. The precision and recall remain the same when BP is used as a classifier.

3. K. Bogahawatte, and S. Adikariin [7] researches have introduced intelligent criminal identification system called ICIS which can potentially distinguish a criminal in accordance with theobservationscollectedfromthecrimelocationforacertainclassofcrimes. The system uses existing evidences in situations for identifying a criminal by clustering mechanism to segment crime data in to subsets, and the Nave Bayesian classification has used for identifying possible suspect of crime incidents.

4. J. Agarwal, R. Nagpal and R. Sehgal in [1] have analyzed crime and considered homicide crime takinginto account the corresponding year and that the trend is descending from 1990 to 2011. They have used the k-means clusteringtechniqueforextractingusefulinformationfromthecrimedatasetusing RapidMinertoolbecauseitissolidandcompletepackage withflexiblesupport options. Figure 2 shows the proposed system architecture.

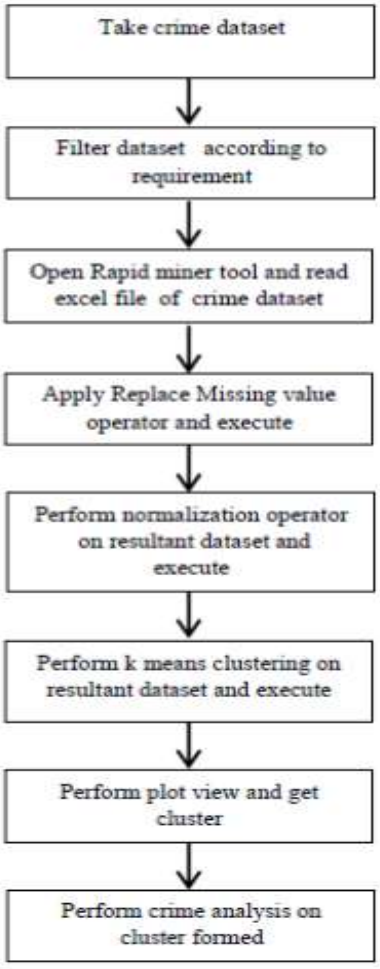

Figure2. Flowchart of crime analysis [1]

5. An integrated system called PrepSearch have proposed by L. Ding et al in [3]. It has been

 ISSN: 2321-8169
124 – 127

combined using two separate categories of visualization tools: providing the geographic view of crimes and visualization ability for social networks. "It will take a given description of a crime, including its location, type, and the physical description of suspects (personal characteristics) as input. To detect suspects, the system will process these inputs through four integrated components: geographic profiling, social network analysis, crime patterns and physical matching". Figure 3 shows the system design and process of PrepSearch.

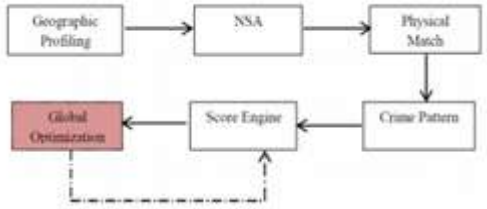

Figure 3. System design and process of PrepSearch.[3]

III. PROPOSEDSYSTEM

The system will look at how to convert crime information into a data-mining problem, so that it will help detectives in solving crimes faster. In terms of crime a cluster is a group of crimes in a geographical region or a hot spot of crime. Whereas, in terms of data mining a cluster is the group of a particular set of objects based on their characteristicsof possiblecrimepattern.
Thusrelevantclustersorasubsetoftheclusterwillhave a one-to-one correspondence to crime patterns. The Proposed system focuses on:

- Crime analysis based on available information to extract crime patterns.
- Using various data mining techniques, frequency of occurring crime can be predicted based on territorial distribution of existing data.
- Crime recognition.

Datamining is needed for crime analysis, because the last is an iterative process of extracting knowledge hidden from large volumes of raw data. To present the proposed model of crime analysis and prediction using datamining, firstwill begin with a big view of this model explained in the following algorithm:

General Algorithm of Proposal Model

Input:Raw data of crime from Government Repository.
Output: Correlated dimensions model for crime analysis and prediction.

Steps:

1. Understandingthecrimedomain: Goals of the crime prediction and detection includes appropriate prior knowledge.
2. Extracting the target dataset:This is for building a dataset for the three dimensions of the proposed model; crime, criminal and geo-crime. By focusing on a subset of variables, feature selection will be done which is not affected by crime conflictions and geo-crime environment changes.
3. Data pre-processing:For mining it is required to improve actual quality of data. The timerequiredforminingthepreprocesseddata is reduced and it also increases mining efficiency. Inourproposalwefocus on data preprocessing to involve data cleaning and treating missing values.
4. Data mining:To introduce correlated patterns AR is applied on each dimension dataset among the three dimensions to advance the crime analysis.
5. Interpretation and Using discovered knowledge:This includesproviding SQL or reports for both separated and correlated dimensions to interpret the discovered patterns. By taking actions based on the knowledge it helps to incorporate this knowledge into the performance system.

A. Algorithm/Flowchart

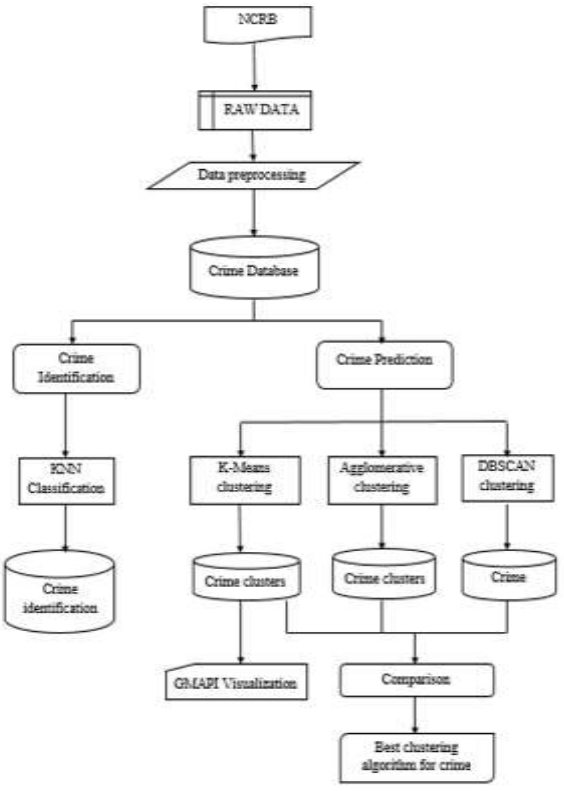

Figure 3.1. Work Flow Diagram [8]

Algorithm: KNNClassification
Input:

1. A finite set D of points to be classified,
2. A finite set T of points,
3. A function c: T –>{1,...,m},
4. A natural number k. Output: A function r: D{1,...,m}

Method:

1. Begin
2. For each x in D do
3. Let U <–{}
4. For each t in T add the pair (d(x,t) , c(t)) to U;
5. Sort the pairs in U using the first components;
6. Count the class labels from the first k elements from U;
7. Let r(x) be the class with the highest number of occurrence;
8. End For each
9. Return r
10. End

Thus proposed system will predict the crime and their patterns of occurrenceso that security can be provided for those areas. It can be applicable for particular region.

IV. SUMMARY

The proposed system presents a new framework for clustering and predicting crimes based on real data. Considering the methods proposed for crime prediction shows that theparameters such as the effect of outliers in the data mining preprocessing, quality of the training and testing data, and the value of features have not been addressed before.TheproposedsystempredictscrimeproneregionsinIndiaon aparticularday.

It will be more accurate if we consider a particular state/region. Also another problem is that the system will not predict the time in which the crime is happening. Since time is an important factor in crime we have to predict not only the crime prone regions but also the proper time.

REFERENCES

[1] ˜J. Agarwal, R. Nagpal, and R. Sehgal, “Crime analysis using k-means clustering,” International Journal of Computer Applications, Vol. 83 No4, December 2013.

[2] J. Han, and M. Kamber, “Data mining: concepts and techniques,” Jim Gray, Series Editor Morgan Kaufmann Publishers, August 2000.

[3] L. Ding et al., “PrepSearch: an integrated crime detection system,” 2009 IEEE 161-163 ISI 2009, June 8-11, 2009, Richardson, TX, USA.

[4] A. Babakura, N. Sulaiman, and M. Yusuf, “Improved method of calssification algorithmsforcrimeprediction,”InternationalSymposiumonBiometricsandSecurity Technologies (ISBAST) IEEE 2014.

[5] S. Sindhiya, and S. Gunasundari, “A survey on Genetic algorithm based feature selection for disease diagnosis system,” IEEE International Conference on Computer Communication and Systems(ICCCS), Feb 2021, 2014, Chermai, INDIA.

[6] S. Sathyadevan, and S. Gangadharan, “Crime analysis and prediction using data mining,” IEEE 2014.

[7] K. Bogahawatte, and S. Adikari, “Intelligent criminal identification system,” IEEE 2013 The 8th International Conference on Computer Science & Education (ICCSE 2013) April 26-28, 2013. Colombo, Sri Lanka.

[8] S.Sivaranjani, Dr.S.Sivakumari and Aasha.M, “Crime Prediction and Forecasting in Tamilnadu using Clustering Approaches ” IEEE 2016.