Comparative Analysis of Crime Rate Prediction using Linear Regression, Logistic Regression and Gradient Boosting Techniques

Miss. Ashwini A. Deshmukh, Prof. M. M. Bartere

Department of Computer Science G. H. Raisoni University, Amravati Amravati,India.
Department of Computer Science G. H. Raisoni University Amravati, India.

Abstract--Urbanization creates a lot of social problems. One of these problems inherent in all cities of the world is crime. Police databases accumulate a large amount of data that could be analyzed in order to reduce crime rates. The analysis of criminal activity and prediction of number of crimes remains one of the most interesting problem for researchers.Crime is one of the most dominant andalarming aspects in the society and its prevention is an importanttask. Crime analysis is a systematic way of detecting andinvestigatingthe patterns and trends that happen in acrime.This paper is focused on analysis of five papers. In this paper, we have studied different crimeprediction techniques and presented the techniques thoroughly.

I. INTRODUCTION

Crime is a violation against the humanity that is oftenaccused and punishable by the law. Criminology is the study ofcrime and it is interdisciplinary sciences that collects andinvestigate data on the crimesand crime performance. The criminalactivities have been increased now-a-days and it is theresponsibility of police and crime investigation department to control and reduce thecrime activities. Crime prediction and criminal identificationare the major problems to the police department as there are large numbers of data of crime exist.Many Algorithms and techniques are used to detect the crime from the databases however Crime is neither systematic nor entirely random. Itebbs and flows with human cycle behaviour, butparticular places are crime attractors. The crime prediction techniques must be able to search for a crime or victim at a faster rate from the huge available records present in the databases of the criminology departments. The techniques and algorithms that are used widely for this purpose have been studied here and taken into account.

II. BACKGROUND

Crime prevention refers to the a series of programs that are involved with individuals, communities, businesses, non-government organizations and all levels of government in addressing the various social and environmental factors that contribute to the risk of community's crime, disorder and victimization. There are many differenttechniques that are available for this purpose. Some tends to be accurate and efficient techniques to predict the crime by many ways. The crime prediction is a very serious issue in today's life. The techniques that are used widely are presented here:

Three Predictive models such as linear regression, logistic regression and gradient boosting are used to predict the crime rate by usingit with Russian Database in [1]. Various Visualization Techniques and Machine Learning Techniques are used for predicting the crime distributed over an area. Raw Databases are processed and visualized based on the need. Afterwards machine learning algorithms are used to extract the knowledge and predict the crime rates[2].A data mining tool named Weka which is an open source tool is used here to analysis the crime. The crime dataset is used here in order to predict the crimes[3].Various Clustering methods are used here to predict the crimes such as K Means, Agglomerative clustering and Density based spatial clustering. The Data Set used for data clustering is collected from NCRB [4]. The Data Mining techniques are used here in order to detect the crime. The dataset is mined in order to find out the crime rate and also to analyse it[5].

The paper is organized as follows: **Section I** Introduction. **Section II** discusses Background. **Section III** discusses previous work. **Section IV** discusses existing methodologies. **Section V** discusses attributes and parameters and how these are affected on mobility models. **Section VI** discusses outcome of result. Finally**Section VII** Conclude this analytical paper.

III. PREVIOUS WORK DONE

The crime rate is increasing day by day. Crime Prediction has become an important part in today's criminal world. The crimes occurring are in higher level and needs to be prevented in order to establish a peaceful society. So various techniques are used in order to find out the crimes that are occurring on a daily basis. The crime prediction helps to predict the crime and reach the criminal faster.The previous work done on this topic is as follows:

VarvaraIngilevichet al. (2018) [1] hascompared different approaches to the problem of forecasting the number of crimes in different areas of the city.The results of model predictions are compared and determined that gradient boosting is the most appropriate method for the problem of crime rate prediction in certain urban area.

Hitesh Kumar Reddy TopiReddyet al.(2018) [2] has provided a framework for visualizing the crime networks and analyzing them by various machine learning algorithms. The project helps the crime analysts to analyze these crime networks by means of various interactive visualizations. FatihahMohdet al.(2017) [3] has presented a comparative study on correlation and information gain algorithms to evaluate and produce the subset of crime features.S. Sivaranjaniet al.(2016) [4] have used various clustering approaches of data mining to analyse the crime data of Tamil Nadu. The crime data is extracted from NationalCrime Records Bureau (NCRB) of India.

Chung-Hsien Yuet al.(2011) [5] has discussed the preliminary results ofa crime forecasting model that is developed in collaboration withthe police department of a United States city in theNortheast.Anna Matsukawa et al. (2018) [6] has surveyed the neighbourhood associations in Kyoto to test a theoreticalmodel for social capital and community based crime prevention.Seiji Shibata et al. (2012) [7] investigated local residentscrime perception about public safety and anxiety related to crimes and attitude towards community crime prevention activities.The result showed that people whohave emergent sense of crisis tend to have an attitude that community crime prevention is their own issuesand willing to participate in them.Masila Abdul Jalil et al. (2017) [8] implemented a prototype called CrimeAnalysis which is based on a framework of ontological-based case matching mechanism. This prototype contains the ontology model and constructed case-matching engine. Suleman Ibrahim et al. (2016) [9] aimed to establish the particularities of cybercrime in Nigeria and whether these suggest problems withprevailing taxonomies of cybercrime. It has examined the explanatory capacity of the existing taxonomies in making sense ofwhat is true of all societies, and what is true of one society at one point in time and space.

IV. EXISTING METHODOLOGIES

Many techniques andschemeshave been implemented over the last several decades.Some techniques seem to be accurate as well as help in the crime detection. The techniques are responsible for predicting the crimes accurately and also to track down the criminals.

A) Predictive Models:Thenumbers of crimes were estimated by using the three models such as linear regression, logistic regression and gradient boosting. The predictive factors used in this model have been selected using the feature selection techniques. This approach allows to increase the accuracy of predictions and to avoid the models over fitting [1].

B) Machine Learning Algorithms:Accurate real-time crime predictions help to reduce the crime rate but remains challenging problem for the scientific community as crime occurrences depend on many complex factors. In this work, various visualizing techniques and machine learning algorithms are adopted for predicting the crime distribution over an area. In the first step, the raw datasets are processed and visualized based on the need. Afterwards, machine learning algorithms were used to extract the knowledge out of these large datasets and discover the hidden relationships among the data which is further used to report and discover the crime patterns that is valuable for crime analysts to analyse these crime networks by the means of various interactive visualizations for crime prediction and hence is supportive in prevention of crimes. [2].

C) Filtering Methods:A comparative study on correlation and information gain algorithm to evaluate and produce the subset of crime features is presented here. The main objective of the study is to find the subset of attributes from a dataset described by a feature set and to classify the crimes into the three categories: low, medium and high. The experiment is carried out using the crime dataset by WEKA, open source data mining software. Based on attributes chosen, the accuracy rates of several classification algorithms were obtained foranalysis. The results proved the accuracy of 96.94%. This method proves to be important in predicting the crimes and also to predict it accurately [3].

D) Clustering Techniques:Clustering approaches of data mining are used here to analyse the crime data of Tamil Nadu. The crime data is extracted from National Crime Records Bureau (NCRB) of India. The data consists of crime information about six cities namely Chennai, Salem,Coimbatore, Madurai, ThiruchirapalliandThirunelvellifrom the year 2000-2014 with 1760 instances and 9 attributes to represent

the instances. Agglomerative clustering, K-Means clustering, and Density Based Spatial Clustering with Noise (DBSCAN) algorithms are used to cluster the crime activities based on some predefined cases and the results of these clustering techniques are compared to find the best suitable clustering algorithm for a specificcrime detection. The result of K-Means clustering algorithm is visualized using Google Map for interactive and easy understanding. The K-NearestNeighbour (KNN) classification is used for crime prediction[4].

E) Data Mining Techniques:Data Mining Techniques are used to detect and predict the crimes. This technique uses the data sets for crime prediction. The datasets contain many counts of crime and crime-related events categorized by the police department. The location and time of these events is saved in the data. Additional spatial and temporal features are extracted from the raw data set and an ensemble of data mining classification techniques is employed to perform the crimeforecasting [5].

V.   ANALYSIS AND DISCUSSION

Crime is classically "unpredictable". It is notrandom, but neither does it take placeconsistently in space or time. A better theoreticalunderstanding is needed to find out the practical crimeprevention solutions that correspond to specific places andtimes.Various types of predictive models are studied and the predictive factors have beenselected by using feature selection techniques. This is done in order to increase the accuracy and efficiency of the models proposed [1]. The visualization technique and machine learning algorithms are adopted for crime prediction technique predicting the crime distribution spread over an area. Raw Data sets are visualized and processed and then machine learning algorithms are applied over to extract the knowledge and predict crimes [2]. Filtering the Data set again can lead to crime prediction and the tool called weka is then used to analyse the crime prediction [3]. The technique proves to be effective in detecting the crimes and also helps the crime departments to investigate about a particular crime and to take the actions accordingly. The clustering techniques are useful in detecting the crimes and also to help the crime investigation department with greater accuracy. Various algorithms are used in order to detect and prevent the crimes with accurate results [4]. The Data Mining technique proves to be effective in order to predict the crimes. The data mining is to be done on raw data set in order to predict the crimes and then different machine learning algorithms are applied onto it in order to predict the crimes [5]. The technique described above has their merits and demerits which are described as below:

TABLE 1: COMPARISONS BETWEEN DIFFERENT SCHEMES.

Proposed scheme and techniques	Advantages	Disadvantages
Predictive Models	The best accuracy was obtained by gradient boosting model.	The proposed system is not implemented yet and negative values can be seen in some models.
Machine Learning Algorithms	Visualizations are provided which tend to be easy to predict the crimes.	The Data used is not up to date.
Filtering Methods	The system has an accuracy rate of 96.94%.	More Algorithms need to be applied for easy prediction.
Clustering Technique	The system proves to be effective for detecting the crimes.	The classification algorithm needs to be improved.
Data Mining Technique	The system is more reliable.	The system is not yet deployed.

VI.   OUTCOME AND POSSIBLE RESULT

Crimes have become a common topic these days and the rate of crimes is increasing day by day. The crime rates need to be taken care of. There are many methods available to detect and predict the crimes and also

to handle it. The techniques prove to be effective in predicting the crime and stop them from occurring. The method has an accuracy rate that is helpful to stop and mitigate the problems of the crimes.

VII. CONCLUSION

This paper is focused on analysis of different techniques and systems such as Predictive Models, Machine Learning Algorithms, Filtering Methods, Clustering Methods and Data Mining Techniques.Thistechnique uses many algorithms and helps to prevent the crimes that occur with on a daily basis. The techniques proves to be helpful to the crime investigation department in order to predict and detect the crimes.

VIII. FUTURE SCOPE

From observations of the methods the future work will include the implementation of a crime prediction system.

REFERENCES

[1]. VarvaraIngilevich,SergeyIvanov,"Crime rate prediction in the urban environment using social factors",Elsevier, 2018.
[2]. Hitesh Kumar Reddy ToppiReddy, BhavnaSaini,GinikaMahajan,"Crime Prediction and Monitoring Framework based on spatial analysis",Elsevier ,2018.
[3]. Masila Abdul Jalil,FatihahMohd, Noor MaizuraMohamad Noor,"A Comparative Study to evaluate Filtering Methods for Crime Data Feature Selection",Elsevier ,2017.
[4]. S.Sivaranjani, Dr. S. Sivakumari, Aasha M., "Crime Detection and Forecasting in Tamil Nadu using Clustering Approaches ",IEEE, 2016.
[5]. Chung-Hsien Yu, Max W. Ward, Melissa Morabito, and Wei Ding, "Crime Forecasting using Data Mining Techniques",IEEE,2011.
[6]. Anna Matsukawa, Shigeo Tatsuki, "Crime prevention through community empowerment: An empiricalstudy of social capital in Kyoto, Japan", Elsevier, 2018.
[7]. Seiji Shibata, Kazunori Hanyu, TatsutoAsakawa,Takahito Shimada & Kenji Omata, "Community Activities to Protect Children from Crime,People's General Trust, and Attitude toward the Activities inJapan", Elsevier, 2018.
[8]. Masila Abdul Jalil, Chia Pui Ling, Noor MaizuraMohamad Noor, FatihahMohd, "Knowledge Representation Model for Crime Analysis", Elsevier, 2017.
[9]. Suleman Ibrahim, "Social and contextual taxonomy of cybercrime:Socioeconomic theory of Nigerian cybercriminals", Elsevier. 2016.