Meenu Rai,2021, 9:1
ISSN (Online): 2348-4098
ISSN (Print): 2395-4752

Design and Implementation of Data Mining Techniques Used In Criminal Activities for Predictive Analysis

M.Tech. Scholar Meenu Rai, Associate Prof. Bhawana Pillai
Dept.of Computer Science & Engg.
Lakshmi Narain College of Technology & Science,
Bhopal, M.P., India

Abstract- In most of the crime detection and hotspot analysis methods, the crime data are processed using different processing techniques to provide accurate analysis. However, there are many problems that lead to poor performance. The major limitations in the literature are the class imbalance problem and multi-objective problems that occur due to the varying classes resulting in the generation of wrong synthetic minority samples. The utilization of graph based clustering methods results in complete graph processing even when the graph has minor changes, thus becoming complicated for further use. The class imbalance problem has to be resolved and the accuracy of the clustering is needed to be enhanced by introducing an efficient integrated sampling and clustering approach for effective serial crime detection. The near-optimal crime detection is needed to be enhanced by not only considering the crime data but also the social crime datasets collected, thus improving the chances of serial crime detection.

Keywords:- Criminal Activities,Predictive Analysis,Data Mining.

I. INTRODUCTION

Data mining is characterized as a knowledge extraction method from large data sets. In other words, data mining is data mining awareness. The vocabulary includes the exploration of results, the language of the query, classification and estimation, decision tree inference, and the study of clusters. Data mining is an interdisciplinary field which combines IT and statistical techniques. Please note that a misnomer is the expression "data mining." It focuses on the identification of data trends and anomalies, but does not contribute to data retrieval itself. The information industry has an enormous amount of data available.

This data are worthless unless they are translated into beneficial data. This vast volume of data needs to be processed and valuable information collected. Data mining also requires such activities such as data purification, data integration, data transformation, data mining,model evaluation and data presentation. Information for a large number of applications like fraud detection, market analysis, control of production, scientific exploration, etc. Although police officers used techniques for predirecting crime for a long time, computer technology has only recently shifted these attempts from basic heuristic methods to complex statistical models, originally focused on local policing experience, using a wide variety of information sources.

These opportunities are used by offenders to participate in illegal acts. Drug traffickers use drug chat rooms. Terrorists and hatred movements use their agendas in social media. Device has an index feature that preserves all communications. If you have access to archived talks on seized machines or on online chat servers, these communications may be useful to the prosecution of crime. Online content may expose participants' lives, social networks, behaviors and preferences However,it is rather difficult and time-consuming to review a vast number of conversations systematically to locate

facts in relation to a court prosecution. Some researchers use standard search techniques to crawl or recover related data in forensic applications or desktop search engines.There are three extraordinary limits to this approach. Current search instruments can detect search-related records and phrases but do not include details on the social networks and actions of a suspect. This easy matching technique is not suited to investigation in cases of crime because the words "drug" and "cocaine" are scarcely used by drug traffickers in their talks. For high-quality search results, the expertise and past knowledge of the search individual may also lead to obsolete, inaccu rate or conflicting details. Search results in many situational fields. In order for current forensic search methods to fix shortcomings.

We are exploring a context in which the classes of individual chat logs are collected and agglomera tions are used to find and summarize the topics of concern in the established communities. Crime inspectors will scan and see the results in the visualizer built. In order to simplify and encourage the investigative process, the aim of this data mining system is to gather instinctive and interpretable evidence from a chat log, particularly in the initial stage where there are not enough indicators for the investigator to start with.

An overview of the framework proposed is prese nted. There are three main system modules: Clique Detector, Idea Miner, and Visualizer of Knowledge. In the chat file, the Clique Detector detects the cliques (communities) It first recognizes all the listed entities in the specified chat log for this reason. The word individual may be the identity of an individual, an organisation, a phone number, or a physical address in the sense of an investigation. For convenience, we assume the organization refers to the name of the person. The Clique Detector uses the co-occurrence frequencies of the entities in chat sessions after the extraction of entities and defines the groups, called cliques.

This chapter includes brief introduction to the crime detection,crimehotspot analysis,detection techniques ,crime mapping concepts, crime analysis techniques, GeographicInformationSystems(GIS)in crime analysis, crime trend prediction using social media and Regre ssion models for crime estimations. The overall crime detection scenarios are analyzed and the detection techniques with their advantages and disadvantages are discussed in detail. The detailed description and evaluation of the various serial crime detection methods are studied for further research. Finally the chapter presents the contribution in the current work and concludes with the organization of the thesis.

II. RELATED WORK

S. Ali, S. A et al[2] In this paper, we take the Punjab province of Pakistan as a case study and apply the proposed model on the real data collected from various police stations. The results show that the identification of potential suspects proved vital for the cases which involves criminals who had previous criminal record. Crimes are considered as social nuisance and results in the degradation of the social values.

M. A. Tayebi, et al[3] propose a probabilistic model of spatial behavior of known offenders within their activity space. Crime Pattern Theory concludes that offenders,rather than venture into unknown territory, frequently commit opportunistic crimes and serial violent crimes by taking advantage of opportunities they encounter in places they are most familiar with as part of their activity space. Our experiments on a large real-world crime dataset show that CRIME TRACER outperforms all other methods used for location recommendation we evaluate here.

The use of data mining techniques can detect the crime patterns and enhance the speed of resolving the crime. Crime clustering is an important process based on data mining approaches (Almanie et al., 2015). Cluster of crimes where there is a lot of crimes in a specified geographical region is referred as geographical group of crime. By using geo-spatial plot of the crime the overlay on the map can be represented visually. To visually locate the hot-spots of crime the densely populated cluster of crime is used. The crime pattern or a crime spree is identified by clustering from a data-mining standpoint. Some widely used crime patterns are the DC sniper, a serial-rapist or a serial killer.These crimes may involve single suspect or may be committed by a group of suspects (Bajpai 2012).For police clustering, the crime dataset is critical material. Any data is kept secret in the dataset,some are available to the public (Shafeeq and Binu 2014).

However, data on narcotics-related offences or juvenile cases is typically more limited. Similarly, in

International Journal of Science, Engineering and Technology

An Open Access Journal

order to alert people in the city, information about sex abusers is made public, but the name of the victim is also avoided. The researcher, a data miner, needs to deal with both these public and private data concerns so that the simulation phase of data mining does not break these ethical boundaries.

The following types of information categories are used in these crime reports: type of crime, date/time, place, etc. Details about the perpetrator (identified or unidentified), the victim and the witness is then presented. In addition, there is the crime story or summary and Modus Operandi (MO) that is typically in the context of the text. Free text is used by police officers or detectives to document much of their findings that can not be found in pre-determined questions in the checkbox.

While the first two types of information are typically stored as binary, character or date table fields in computer databases, the third one is mostly stored as free text. The problem of data mining crime data also emerges from the field of free text. While free text fields can provide a perfect story line for the newspaper columnist, translating them into data mining attributes is not always an easy task (Tayal et al., 2015).

In data mining, the role of the clustering algorithm (Gera and Vohra 2014) is to classify related record classes, but separate from the rest of the data. By using these clusters, a criminal spree committed by one or the same group of offenders is identified. The next boring technique with this knowledge is to identify the variables that have the best clustering. The detectives would then be confronted with these clusters to dig down using their domain information.

Next, the police work on the spree of crime and tackle one of the murders that can be carried out by automatic crime pattern identification. The automatic identification of the trend of crime helps in the resolution of the entire spree or, in some situations, where the clusters of events are accused of being one spree, it is possible to create full facts from the various bits of details from each of the crime incidents (Sharma and Kumar 2013).

III. PROPOSED METHODOLOGY

Most technologies on which data mining is based are based on statistics,such as regression analysis, standard distribution, standard deviation, standard variance, discrimination analysis, cluster analysis, and confidence intervals.To research data and data interactions, both of these are included.
Artificial Intelligence Artificial intelligence, or AI, which is based on heuristics as opposed to statistics, seeks to apply mathematical problems to human thought-like computation. Such AI principles that have been embraced by some high-end commercial products, such as Relational Database Management Systems query optimization modules (RDBMS).

Machine Learning The union of statistics and AI is machine learning. As it combines AI heuristics with sophisticated statistical analysis, it may be called an evolution of AI.

Machine learning aims to help computer programs learn about the data they are analyzing, such that programs make various choices depending on the features of the data analyzed, use statistics for simple principles, and incorporate more complex AI heuristics and algorithms to accomplish their goals. Study Using Denver Neighborhood Demographics Dataset, we applied some population analysis after completing our main objective by identifying geographical and temporal criminal hotspots and forecasting possible forms of crime. By analyzing the correlation between the crime rate in each neighborhood and its description of demographic statistics, we decided to better explain our model outcomes. We find that the distribution of age and gender between populations varied between unsafe and healthy areas.

International Journal of Science, Engineering and Technology

An Open Access Journal

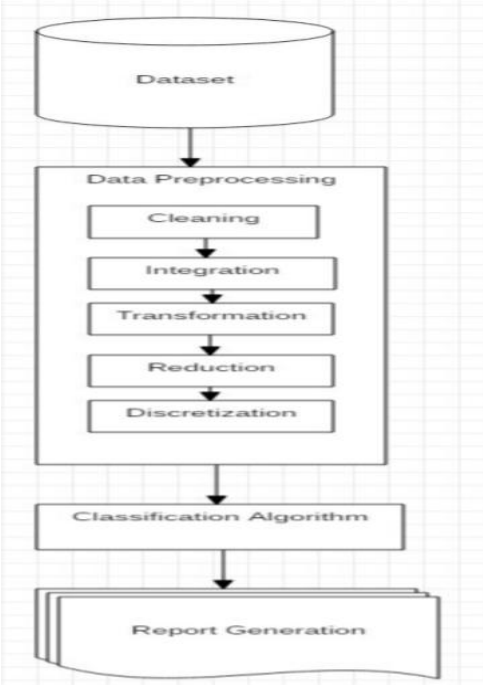

Fig 1. Data classification process using data mining technique.

In particular, unsafe communities have more males while the safe neighborhoods have more females We have not found any link between hotspots of crime and distribution of race among peoples.

Social crime data set and the gathered dataset from police stations are considered in this work for predicting the locations in which serial crimes are happening most in the real world environment. This data set is gathered from the multiple social media web sites in terms of different crimes activities happening in the different crime locations. This data set would consists of the details like crime type, number of crimes happening in given time period, people opinion about those particular crime happened in different location. By using these information, the different types of crimes are analyzed and finally the similar types of crimes are grouped together to predict the various crimes

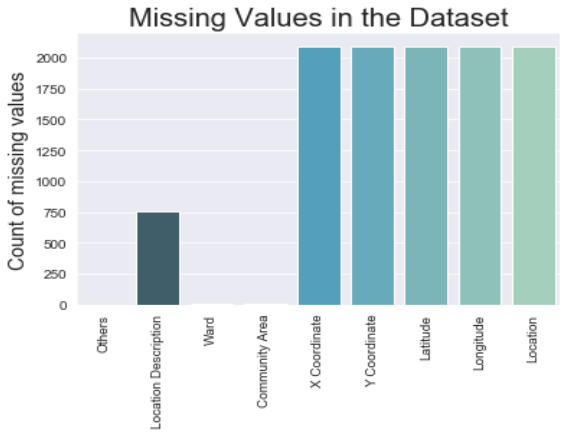

Fig 2. Count the missing values of crime data sets.

IV. CONCLUSION

The second goal of our analysis was to predict the type of crime that might occur in a given area during a certain time frame.The Bayesian classifier has made it possible for us to achieve this objective with fair precision. You ought to have four associated charact eristics of the crime to predict an upcoming form of crime. The needed attributes are: the month of occurrence, the day of the week of occurrence, the time of occurrence, and the position of the crime. All the attributes can be viewed in their nominal values.

REFERENCE

[1] M.J.C. Baculo, C.S. Marzan, R.de Dios Bulos and C. Ruiz,"Geospatial-temporal analysis and classifi cation of criminal data in Manila," 2017 2nd IEEE International Conference on Computational Inte lligence and Applications (ICCIA), Beijing, 2017, pp.6-11, doi: 10.1109/CIAPP.2017.8167050.

[2] S.Ali, S.A. Alvi and A.U. Rehman,"The Usual Suspects:Machine Learning Based Predictive Policing for Criminal Identification,"2019 13th International Conference on Open Source Syste ms and Technologies (ICOSST), Lahore, Pakistan, 2019,pp.16,doi:10.1109/ICOSST48232.2019.9043 925.

[3] M.A.Tayebi, M. Ester, U. Glässer and P.L. Brantin gham,"CRIMETRACER: Activity space based crime location prediction,"2014 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2014), Beijing, 2014,pp.472480,doi:10.1109/ASONAM.2014.6921 628.

[4] X.Wang, D.E. Brown and M.S. Gerber,"Spatio-temporal modeling of criminal incidents using

International Journal of Science, Engineering and Technology

An Open Access Journal

geographic, demographic, and twitter-derived information," 2012 IEEE International Conference on Intelligence and Security Informatics, Arlington,VA,2012,pp.3641,doi:10.1109/ISI.2012.6284088.

[5] N.Mahmud, K.I. Zinnah, Y.A. Rahman and N. Ahmed,"Crimecast:A crime prediction and strategy direction service,"2016 19th International Conference on Computer and Information Technology (ICCIT), Dhaka, 2016, pp. 414-418, doi: 10.1109/ICCITECHN.2016.7860234.

[6] L. Elluri, V. Mandalapu and N. Roy, "Developing Machine Learning Based Predictive Models for Smart Policing,"2019 IEEE International Conference on Smart Computing (SMARTCOMP), Washington, DC, USA, 2019, pp.198-204, doi: 10.1109/SMARTCOMP.2019.00053.

[7] S. K. Dash, I. Safro and R. S. Srinivasamurthy, "Spatio-temporal prediction of crimes using network analytic approach," 2018 IEEE International Conference on Big Data (Big Data), Seattle, WA, USA, 2018, pp. 1912-1917, doi: 10.1109/BigData.2018.8622041.

[8] J. Hughes, B. Collier and A. Hutchings, "From playing games to committing crimes: A multi-technique approach to predicting key actors on an online gaming forum," 2019 APWG Symposium on Electronic Crime Research (eCrime), Pittsburgh, PA, USA, 2019, pp.1-12, doi: 10.1109/eCrime47957.2019.9037586.

[9] R. K. Vishwakarma and R. Shankar, "Modeling brain and behavior of a terrorist through fuzzy logic and ontology," 2013 IEEE International Conference on Industrial Engineering and Engineering Management, Bangkok, 2013, pp. 857-861, doi: 10.1109/IEEM.2013.6962533.

[10] Q. A. Al-Haija and L. Tawalbeh, "Autoregressive Modeling and Prediction of Annual Worldwide Cybercrimes for Cloud Environments," 2019 10th International Conference on Information and Communication Systems (ICICS), Irbid, Jordan, 2019, pp.47-51, doi: 10.1109/IACS.2019.8809125.

[11] A. Singh Rajawat and S. Jain, "Fusion Deep Learning Based on Back Propagation Neural Network for Personalization,"2nd International Conference on Data, Engineering and Applications (IDEA), Bhopal, India, 2020, pp. 1-7, doi: 10.1109/IDEA49133.2020.9170693.

[12] J. Azeez and D. J. Aravindhar, "Hybrid approach to crime prediction using deep learning," 2015 International Conference on Advances in Computing, Communications and Informatics (ICACCI), Kochi, 2015, pp. 1701-1710, doi: 10.1109/ICACCI.2015.7275858.

[13] S.S. Khan and J.Wei, "Real-Time Power Outage Detection System using Social Sensing and Neural Networks," 2018 IEEE Global Conference on Signal and Information Processing (Global SIP), Anaheim, CA, USA, 2018, pp. 927-931, doi: 10.1109/GlobalSIP.2018.8646443.