A REVIEW OF DETECTION OF CRIME USING DATA CLASSIFICATION

Shivneet Singh
Email id :- redhu.shivneet3@gmail.com

Abstract— **When crimes occur regularly in a society, they will have an impact on the institutions and organizations there. Because of this, it is important to investigate the causes of crime, as well as the variables and relationships that contribute to its occurrence, in order to discover the best methods for managing crime and preventing its spread. The major purpose of this research is to categorize clustered crimes according to their recurrence frequency throughout various years. Many types of criminal activity are analyzed, investigated, and pattern-discovered via data mining. To an actual crime dataset collected by the police in England and Wales between 1990 and 2011, we used a theoretical model based on data mining methods like clustering and classification. To enhance the quality of the model and get rid of aspects that weren't adding anything to it, we gave each one a weight. Outlier Detection operator settings are optimized using the RapidMiner software and the Genetic Algorithm (GA).**

Keywords— Crime information report, ,statistics data analysis; Clustering; K-Means Algorithm;

I. INTRODUCTION

A. Crime Analysis

Security services nowadays cannot function without routinely collecting and analyzing data connected to criminal activity. The most crucial issues to be resolved include the use of a consistent approach to categorize these data according to the frequency and place of occurrence, the discovery of the underlying pattern among the crimes committed at various periods, and the forecast of their future association.

Using actual datasets and presenting an appropriate framework that is immune to outliers are both important factors to think about in this context. When it comes to data mining, outliers may have a major impact on the outcomes during the preprocessing step. Therefore, a mechanism should be found to discover the outlier data and then delete them. Using Accuracy and Classification error, we define a Fitness function and use it to optimize the parameters of the Outlier Detection operator using the GA. Due to their detrimental effects on data clustering and classification, low-value features were weeded out using a weighting approach, leading to improved prediction accuracy and reduced classification error..

The main purposes of crime analysis are mentioned below [1]:

- Extraction of crime patterns by crime analysis and based on available criminal information,
- Prediction of crimes based on spatial distribution of existing

other variable.‖ Figure 2 shows the proposed system architecture.

In [12] an integrated system called PrepSearch have proposed by G. Jiji-S. Anantharadha. It has been 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

data and prediction of crime frequency using various data mining techniques,

B. Crime recognition. *Clustering*

It is termed "clustering" when a group of data or objects is broken up into smaller groups. That is, a cluster is made up of related data that exhibits the same characteristics as a whole. Clustering is similar to classification in that both involve grouping data into categories, however unlike classification, the classes in clustering are not predetermined and the grouping is done automatically [2].

C. Clustering by K-means Algorithm

Of the many clustering algorithms used in scientific and commercial applications, K-means is by far the most popular [3] [4] [5] due to its simplicity and widespread use. Because of its ease of use, the K-means has gained widespread popularity. Although its computational cost rises linearly with the number of data points, this approach is nevertheless well-suited for clustering huge datasets [5]. The K-means algorithm's efficiency may be diminished by a variety of factors, including the user's need to choose the number of clusters, the impact of outlier data, the high dimensionality of the data, and the algorithm's sensitivity toward centers for initial clusters [6].

II. LETRATURE REVIEW

MINGCHEN FENG , JIANGBIN ZHENG [1] 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 because it is solid and complete package with flexible support options. Figure1 shows the proposed system architecture.

Deepika K.K, Smitha Vinod in [2] 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, robberyand etc. Their work will be helpful for the local police stations in decision making and crime supervision.

―After training systems will predict data values for next coming fifteen years. The system is trained by applying linear regression over previous year data. This will produce a formula and squared correlation( ).

The formula is used to predict values for comong future years. The coeffecent of determination, , is useful because is gives the proportion of variance of one variable that is predictable from

four integrated components: geographic profiling, social network analysis, crime patterns and physical matching.‖Figure 3 shows the system design and process of PrepSearch.
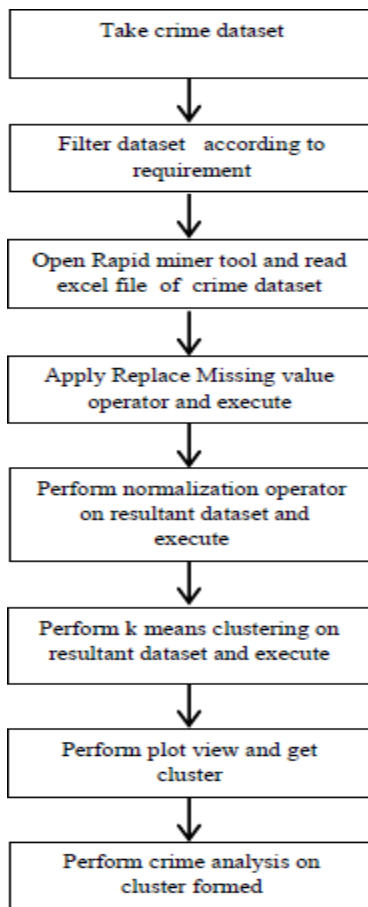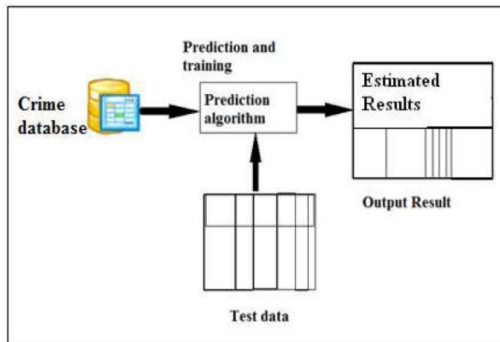
Fig. 1. Flow chart of crime analysis [1]

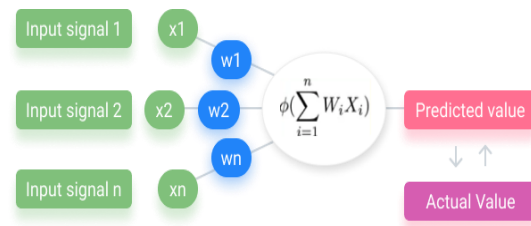Fig. 2. Predicting future crime trends [11]

Fig. 3. System design and process of PrepSearch [12]

In [13] researchers have proposed intelligent criminal identification system called ICIS which may possibly recognize a criminal in accordance with the observations acquired from the crime place for a given class of crimes. Clustering method is utilized to divide crime data into subgroups, and Naive Bayesian classification is used to find potential suspects in criminal cases. ICIS has harnessed the communication prowess of a multi agent system to speed up the process of identifying suspects. ICIS is broken down into its component parts—the user interface, managed bean, multi agent system, and database—so that its architecture can be described. Database implementation uses Oracle Database, and the Java platform is utilized for both criminal pattern detection and pattern implementation.

A. Babakura, N. Sulaiman, and M. Yusuf offer a more effective approach of categorization algorithms for crime prediction in [14]. They analyzed the accuracy of crime prediction using Naive Bayesian and Back Propagation (BP) classification algorithms for each of the 50 states in the USA. The training is used to construct the model, which is then put to use in the second stage. The performance parameters such as Accuracy, Precision and Recall are utilized for comparison of the classification algorithms. When using BP as a classifier, performance metrics like as accuracy and recall are unaffected.

Using data mining, researchers have offered crime analysis and prediction [15]. They have developed a method that combines computer science with criminal justice to create a data mining process that may aid in the swift resolution of criminal cases. Also they have concentrated on reasons of crime occurrence such criminal past of offender, politics, animosity and crime variables of each day. Data is gathered, categorized, patterns are identified, predictions are made, and results are shown as part of their methodology.

III. EXISTING SYSTEM

Crime control and crime suppression are the two primary areas where data mining in criminology may be used. Using a novel distance metric for comparing all people based on their characteristics and then grouping them appropriately, De Bruin et al. [1] presented a framework for crime trends. When it comes to e-governance projects, Manish Gupta et al. [2] both shine a light on the current systems used by Indian police and offer an interactive query based interface as a crime analysis tool to aid police in their operations. The suggested interface is used to glean relevant data from NCRB's massive crime database and identify

problem areas.

methods of data mining, such clustering and other similar processes. The suggested interface has been tested and shown to reduce crime in India.

Architecture

1. THE SUGGESTIONS FOR THE SYSTEM

records. Nazlena Mohamad Ali et al.[3] outline the methodology, user surveys, the system design, and the future plans for the Visual Interactive Malaysia Crime News Retrieval System (i-JEN). The study's primary goals were to build crime-based events, explore how these events can be used to enhance classification and clustering, create a system for retrieving crime news that is both effective and interactive, integrate these features into a usable, high-performing whole, and conduct an evaluation of the system's usability and performance.

Discrepancy detection in audit is one area where Sutapat Thiprungsri [4] looks into the use of cluster analysis in the accounting field. The focus of his research is on how clustering may be used in audits to automate fraud screening. When auditing claims for group life insurance, he employed cluster analysis to assist auditors zero in on the most important information. In order to forecast criminal trends and save investigation times, A. Malathi et al.[5] investigate a data mining strategy based on missing value and clustering algorithms. To foresee future criminal behavior, Malathi. A et al.[6] used a clustering/classify based approach. Crime statistics collected by the city's police department are analyzed using data mining methods. These data mining efforts may provide insights that may be utilized to reduce and perhaps prevent crime in the years to come. Dr. S. Santhosh Baboo and Malathi conducted research to create a crime analysis tool for the Indian context by combining many data mining methods.
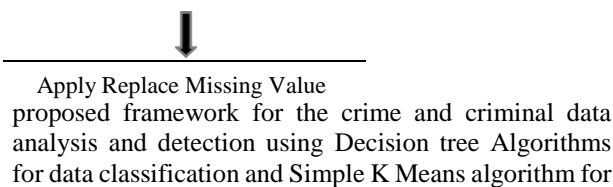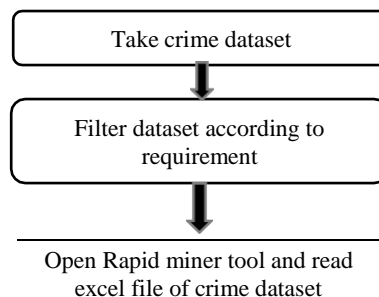
efficiently handle crime investigation.

The proposed tool enables agencies to easily and economically clean, characterize and analyze crime data to identify actionable patterns and trends .Kadhim B. Swadi Al-Janabi [8] presents a

After literature review there is need to used an open source data mining tool which can be implemented easily and analysiscan be done easily. So here crime analysis is done on crime dataset by applying k means clustering algorithm using rapid miner tool.

The procedure is given below:

1. First we take crime dataset
2. Filter dataset according to requirement and create new dataset which has attribute according to analysis to be done
3. Open rapid miner tool and read excel file of crime dataset and apply "Replace Missing value operator" on it and execute operation
4. Perform "Normalize operator" on resultant dataset and execute operation
5. Perform k means clustering on resultant dataset formed after normalization and execute operation
6. From plot view of result plot data between crimes and get required cluster
7. Analysis can be done on cluster formed.

Take crime dataset

Filter dataset according to requirement

Open Rapid miner tool and read excel file of crime dataset

Apply Replace Missing Value

proposed framework for the crime and criminal data analysis and detection using Decision tree Algorithms for data classification and Simple K Means algorithm for

data clustering. The paper tends to help specialists in discovering patterns and trends, making forecasts, finding relationships and possible explanations, mapping criminal networks and identifying possible suspects. Aravindan Mahendiran et al. [9] apply myriad of tools on crime data sets to mine for information that is hidden from human perception. With the help of state of the art visualization techniques we present the patterns discovered through our algorithms in a neat and intuitive way that enables law enforcement departments to channelize their resources accordingly. Sutapat Thiprungsri[10] examine the possibility of using clustering technology for auditing. Automating fraud filtering can be of great value to continuous audits. The objective of their study is to examine the use of cluster analysis as an alternative and innovative anomaly detection technique in the wire transfer system. K. Zakir Hussain et al. [11] tried try to capture years of human experience into computer models via data mining and by designing a simulation model.
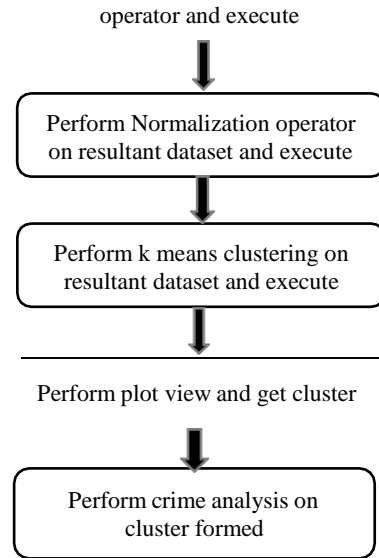
operator and execute

Perform Normalization operator on resultant dataset and execute

Perform k means clustering on resultant dataset and execute

Perform plot view and get cluster

Perform crime analysis on cluster formed

Fig 1: Flow chart of crime statistical analysis

IV. CONCLUSION AND FUTURE SCOPE

Future scope of CIR includes the implementation of online Data Base which will be useful to implement when the project is implemented fully. The Application need to be platform independent.

Useful for transforming the raw data into meaningful and crucial information which will be helpful in forming Decision Clusters for future use. If the number of crime cases are extreme, extra security can be provided to the society with additional security forces which helps for investigation agencies.

V. REFERENCES

[1] . MINGCHEN FENG , JIANGBIN ZHENG, JINCHANG REN , (Senior Member, IEEE), AMIR HUSSAIN , (Senior Member, IEEE), XIUXIU LI, YUE XI , AND QIAOYUAN LIU on Big Data Analytics and Mining for Effective Visualization and Trends Forecasting of Crime Data Received May 8, 2019, accepted July 9, 2019, date of publication July 22, 2019 *Digital Object Identifier 10.1109/ACCESS.2019.2930410.*

[2] Deepika K.K, Smitha Vinod on "Crime analysis in India using data mining techniques" INTERNATIONAL JOURNAL OF ENGINEEEING & TECHNOLOGY

[3] . International Journal of Advanced Computational Engineering and Networking, ISSN: 2320-2106, Volume-4, Issue-5, May.-

[4]

[5] J. Han, and M. Kamber, —Data mining: concepts and techniques,‖ 2nd Edition, Morgan Kaufmann Publisher, 2001.

[6] S. Joshi, and B. Nigam, —Categorizing the document using multi class classification in data mining,‖ International Conference on Computational Intelligence and Communication Systems, 2011.

[7] T. Phyu, —Survey of classification techniques in data mining,‖ Proceedings of the International Multi Conference of Engineers and Computer Scientists Vol. IIMECS 2009, March 18 - 20, 2009, Hong Kong.

[8] S.B. Kim, H.C. Rim, D.S. Yook, and H.S. Lim, —Effective Methods for Improving Naïve Bayes Text Classifiers,‖ In Proceeding of the 7th Pacific Rim International Conference on Artificial Intelligence, Vol.2417, 2002.

[9] 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 20-21, 2014, Chermai, INDIA.

[10] P. Gera, and R. Vohra, —Predicting Future Trends in City Crime Using Linear Regression,‖ IJCSMS (International Journal of Computer Science & Management Studies) Vol. 14, Issue 07Publishing Month: July 2014.

[11] L. Ding et al., —PerpSearch: an integrated crime detection system,‖ 2009 IEEE 161-163 ISI 2009, June 8-11, 2009, Richardson, TX, USA.

[12] 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.

[13] A. Babakura, N. Sulaiman, and M. Yusuf, —Improved method of calssification algorithms for crime prediction,‖ International Symposium on Biometrics and Security Technologies (ISBAST) IEEE 2014. [15] S. Sathyadevan, and S. Gangadharan, —Crime analysis and prediction using data mining,‖ IEEE 2014.