Crime Analysis Using K-Means Clustering

[1]Khushabu A. Bokde, [2]Tiksha P. Kakade,
[3]Dnyaneshwari S. Tumsare, [4]Chetan G. Wadhai
B.E. Student
Department of CSE, Ballarpur Institute of Technology,
Ballarpur,Chandrapur District, Maharashtra, India.

Prof. Deepa Bhattacharya
Assistant Professor
Department of CSE,
Ballarpur Institute of Technology

Abstract - **In today's world security is an aspect which is given higher priority by all political and government worldwide and aiming to reduce crime incidence. As data mining is the appropriate field to apply on high volume crime dataset and knowledge gained from data mining approaches will be useful and support police force. So In this paper crime analysis is done by performing k-means clustering on crime dataset using rapid miner tool and deploy on the web server.**

Keywords— Crime; Clustering; K-Means Algorithm;

1. INTRODUCTION

In present scenario criminals are becoming technologically sophisticated in committing crime and one challenge faced by intelligence and law enforcement agencies is difficulty in analyzing large volume of data involved in crime and terrorist activities therefore agencies need to know technique to catch criminal and remain ahead in the eternal race between the criminals and the law enforcement. So appropriate field need to chooses to perform crime analysis and as data mining refers to extracting or mining knowledge from large amounts of data, data mining is used here on high volume crime dataset and knowledge gained from data mining approaches is useful and support police forces. To perform crime analysis appropriate data mining approach need to be chosen and as clustering is an approach of data mining which groups a set of objects in such a way that object in the same group are more similar than those in other groups and involved various algorithms that differ significantly in their notion of what constitutes a cluster and how to efficiently find them. In this paper k means clustering technique of data mining used to extract useful information from the high volume crime dataset and to interpret the data which assist police in identify and analyze crime patterns to reduce further occurrences of similar incidence and provide information to reduce the crime. In this paper k mean clustering is implemented using open source data mining tool which are analytical tools used for analyzing data .Among the available open source data mining suite such as R, Tanagra ,WEKA ,KNIME ,ORANGE ,Rapid miner.k means clustering is done with the help of rapid miner tool which is an open source statistical and data mining package written in Java with flexible data mining support options. Also for crime analysis dataset used is Crime dataset an offences recorded by the police in India.

This paper is divided into 7 sections: Related work, Proposed System Architecture, Experimental set up & Results, Conclusion, Future scope, References

1.1 Crime analysis

Crime analysis is defined as analytical processes which provides relevant information relative to crime patterns and
trend correlations to assist personnel in planning the deployment of resources for the prevention and suppression of criminal activities

It is important to analyze crime due to following reasons :

1. Analyze crime to inform law enforcers about general and specific crime trends in timely manner

2. Analyze crime to take advantage of the plenty of information existing in justice system and public domain.

Crime rates are rapidly changing and improved analysis finds hidden patterns of crime, if any, without any explicit prior knowledge of these patterns.

The main objectives of crime analysis include:

1. Extraction of crime patterns by analysis of available crime and criminal data
2. Prediction of crime based on spatial distribution of existing data and anticipation of crime rate using different data mining techniques
3. Detection of crime

2. RELATED WORK

Data mining in the study and analysis of criminology can be categorized into main areas, crime control and crime suppression. De Bruin et. al. [1] introduced a framework for crime trends using a new distance measure for comparing all individuals based on their profiles and then clustering them accordingly. Manish Gupta et. al. [2]. highlights the existing systems used by Indian police as e-governance initiatives and also proposes an interactive query based interface as crime analysis tool to assist police in their activities. He proposed interface which is used to extract useful information from the vast crime database maintained by National Crime Record Bureau (NCRB) and find crime hot spots using crime data mining techniques such as clustering etc. The effectiveness of the proposed interface has been illustrated on Indian crime records. Nazlena Mohamad Ali et al.[3] discuss on a development of Visual Interactive Malaysia Crime News Retrieval System (i-JEN) and describe the approach, user studies and planned, the system architecture and future plan. Their main objectives were to construct crime-based event; investigate the use of crime based event in improving the classification and clustering; develop an interactive crime news retrieval system; visualize crime news in an effective and interactive way; integrate them into a usable and robust system and evaluate the usability and system performance and the study will contribute to the better understanding of the crime data consumption in the Malaysian context as well as the developed system with the visualization features to address crime data and the eventual goal of combating the crimes .Sutapat Thiprungsri [4] examines the application of cluster analysis in the accounting domain, particularly discrepancy detection in audit. The purpose of his study is to examine the use of clustering technology to automate fraud filtering during an audit. He used cluster analysis to help auditors focus their efforts when evaluating group life insurance claims. A. Malathi et al.[5] look at the use of missing value and clustering algorithm for a data mining approach to help predict the crimes patterns and fast up the process of solving crime. Malathi. A et. al.[6] used a clustering/classify based model to anticipate crime trends. The data mining techniques are used to analyze the city crime data from Police Department. The results of this data mining could

potentially be used to lessen and even prevent crime for the forth coming years.Dr. S. Santhosh Baboo and Malathi. A [7] research work focused on developing a crime analysis tool for Indian scenario using different data mining techniques that can help law enforcement department to 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 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.

3. PROPOSED SYSTEM ARCHITECTURE

After literature review there is need to used an open source data mining tool which can be implemented easily and analysis can be done easily. So here crime analysis is done on crime dataset by applying k means clustering algorithm using rapid miner tool and that result to be stored on web servers and shows the brief ideas about the analysis

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
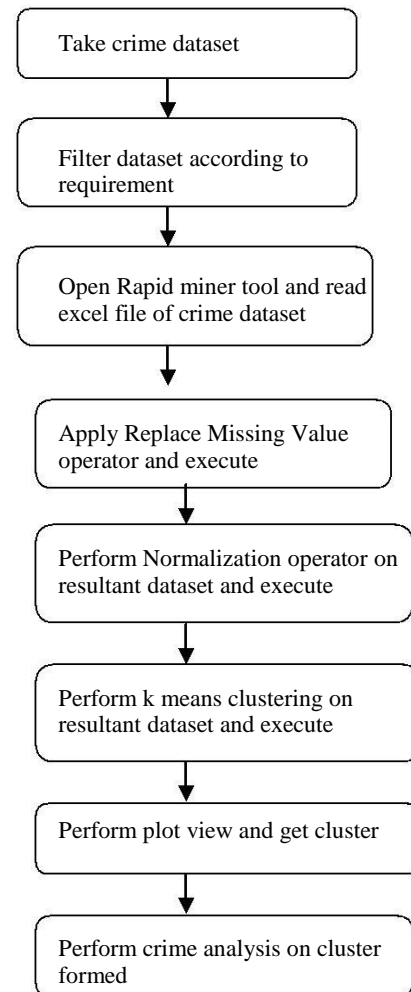7. Analysis can be done on cluster formed.

Fig 1: Flow chart of crime analysis

4. EXPERIMENTAL SETUP AND RESULTS

4.1 Approach Used

4.1.1 k-means algorithm

K-means clustering is one of the method of cluster analysis which aims to partition n observations into k clusters in which each observation belongs to the cluster with the nearest mean.

Process

1. Initially, the number of clusters must be known let it be k
2. The initial step is the choose a set of K instances as centres of the clusters.
3. Next, the algorithm considers each instance and assigns it to the cluster which is closest.
4. The cluster centroids are recalculated either after whole cycle of re-assignment or each instance assignment.
5. This process is iterated.

K means algorithm complexity is O(tkn), where n is instances, c is clusters, and t is iterations and relatively efficient . It often terminates at a local optimum. Its disadvantage is applicable only when mean is defined and need to specify c, the number of clusters, in advance. It unable to handle noisy data and outliers and not suitable to discover clusters with non-convex shapes.

Published by :

http://www.ijert.org

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 7 Issue 04, April-2018

4.2 Dataset Used

Crime dataset used for crime analysis is an offences recorded by the police in india by offence and police force area from 2013 to 2017-18 [12].In Table 1 sample crime dataset is shown.

Table 1. Crime dataset

Year	Murder	Roberry	Kidnap	dowary
2013	10	19	15	20
2014	6	56	23	21
2015	6	54	24	56
2016	6	29	17	45
2017-2018	20	28	32	35

4.3 Tool Used

Many open source data mining suites are available such as R, Tanagra, Weka , KNIME, Orange, Rapid miner. Here we are performing crime analysis using Rapid miner tool because of following reason:

1. It is solid and complete package with Flexible/affordable support options.
2. Enterprise-ready performance and scalability for big data analytics Innovative analyst support
3. We can program by piping components together in a graphic ETL work flows.

Also it has good features that if you set up an illegal work flows Rapid Miner suggest Quick Fixes to make it legal.

4.4. K means cluster analysis

This involves tracking crime rate changes from one year to the next and used data mining to project those changes into the future. Here we consider crime and plot it with year and analysis variation in graph on cluster formed.
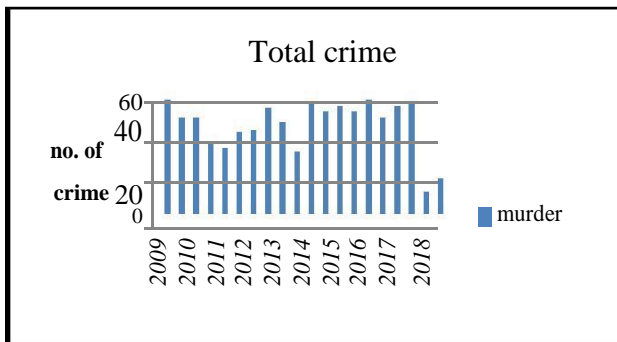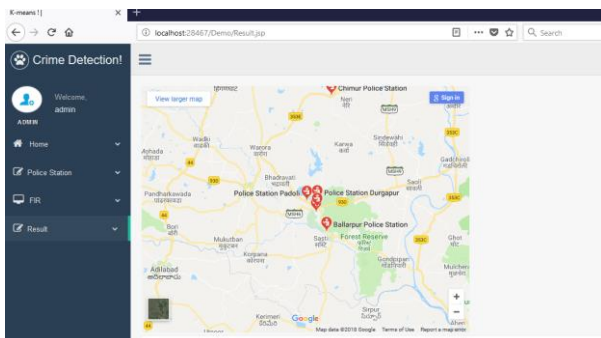
Fig 1: total number of crime 200 9-2018

Output 1:

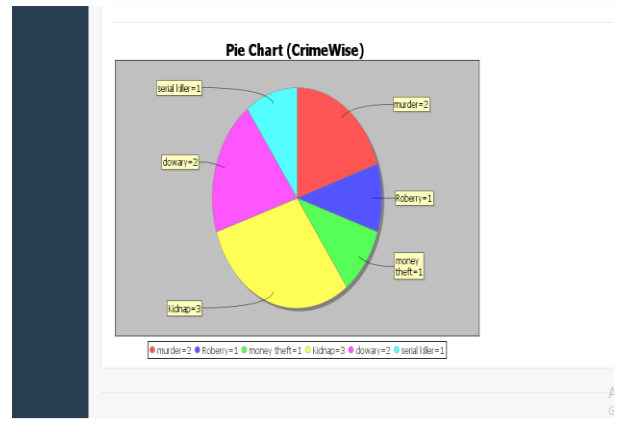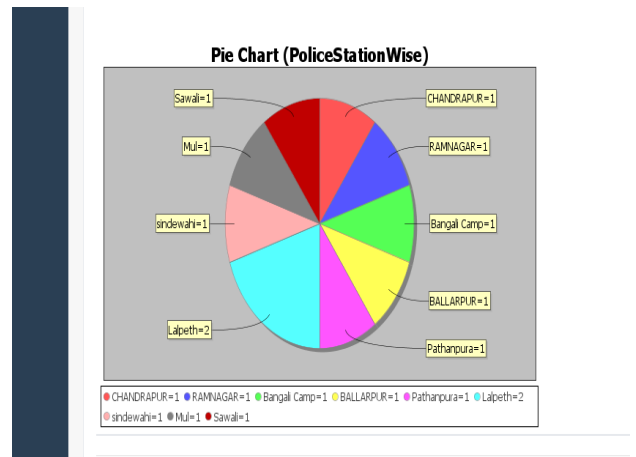Fig 2 Total no.of police station used in dataset

Output 2:

Fig 3 Total no.of crime during the year 2009-2011 using piechart

Output 3:

From Fig 5 it can be seen that the adding the number of criminal records in to the the dataset for analysis.
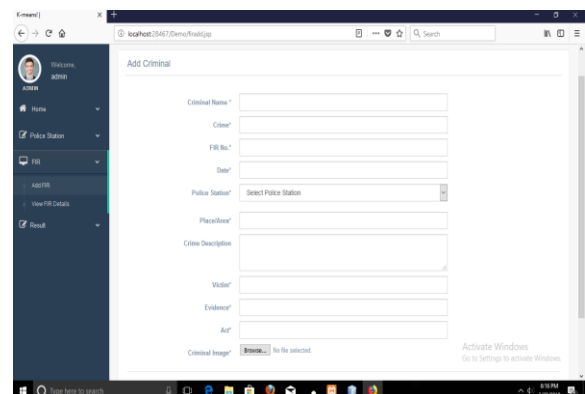
Fig 5: adding criminal information

5. CONCLUSION

This project focuses on crime analysis by implementing clustering algorithm on crime dataset using rapid miner tool and here we do crime analysis by considering crime homicide and plotting it with respect to year and got into conclusion that homicide is decreasing from 2009 to 2018 .From the clustered results it is easy to identify crime trend over years and can be used to design precaution methods for future.

6. FUTURE SCOPE

From the encouraging results, we believe that crime data mining has a promising future for increasing the effectiveness and efficiency of criminal and intelligence analysis. Visual and intuitive criminal and intelligence investigation techniques can be developed for crime pattern. As we have applied clustering technique of data mining for crime analysis we can also perform other techniques of data mining such as classification. Also we can perform analysis on various dataset such as enterprise survey dataset, poverty dataset, aid effectiveness dataset, etc.

7. REFERENCES

[1] De Bruin ,J.S.,Cocx,T.K,Kosters,W.A.,Laros,J. and Kok,J.N(2006) Data mining approaches to criminal carrer analysis ,"in Proceedings of the Sixth International Conference on Data Mining (ICDM"06) ,Pp. 171-177

[2] Manish Gupta1*, B.Chandra1 and M. P. Gupta1,2007 Crime Data Mining for Indian Police Information System

[3] Nazlena Mohamad Ali1, Masnizah Mohd2, Hyowon Lee3, Alan F. Smeaton3, Fabio Crestani4 and Shahrul Azman Mohd Noah2 ,2010 Visual Interactive Malaysia Crime News Retrieval System

[4] Sutapat Thirprungsri Rutgers University .USA ,2011 Cluster Analysis of Anomaly Detection in Accounting Data : An Audit Approach 1

[5] A.Malathi ,Dr.S.Santhosh Baboo. D.G. Vaishnav College,Chennai ,2011 Algorithmic Crime Prediction Model Based on the Analysis of Crime Clusters.

[6] Malathi.A 1 ,Dr.S.Santhosh Baboo 2 and Anbarasi . A 31 Assistant professor ,Department of Computer Science ,Govt Arts College ,Coimbatore , India . 2 Readers , Department of Computer science , D.G. Vaishnav Collge ,Chennai , India , 2011 An intelligent Analysis of a city Crime Data Using Data Mining

[7] Malathi , A; Santhosh Baboo , S, 2011 An Enhanced Algorithm to Predict a Future Crime using Data Mining

[8] Kadhim B.Swadi al-Janabi . Department of Computer Science . Faculty of Mathematics and Computer Science

    University of Kufa/Iraq , 2011 A Proposed Framework for Analyzing Crime DataSet using Decision Tree and Simple K-means Mining Algorithms.

[9] Aravindan Mahendiran, Michael Shuffett, Sathappan Muthiah, Rimy Malla, Gaoqiang Zhang,2011 Forecasting Crime Incidents using Cluster Analysis and Bayesian Belief Networks

[10] Sutapat Thiprungsri,2012 Cluster Analysis for Anomaly Detection in Accounting Data : An Audit Approach1