CRIME ANALYSIS AND PREDICTION USING K-MEANS CLUSTERING TECHNIQUE

[1]Wasim A. Ali
Research Scholar, Department of Computer Engineering, VTU-RC, Mysore, Karnataka, India.

[2] Husam Alalloush
PG. Department of Computer Science, St.Pholimena's College, Mysore, Karnataka. India.

[3]Manasa K.N.
Research scholar, PET Research Center(affiliated to University of Mysore), PES college of engineering, Mandya, Karnataka,India

ABSTRACT
Analysis of crime is a collection of strategies that allow the police forces to become more effective through better knowledge. Our proposed framework aims to forecast the probability of a crime occurring in a city by analyzing the crime dataset and visualizing the findings for better comprehension on a google map. Here we can predict most places where the offense will happen. This research is achieved by using the clustering algorithm of K-means that group related objects into clusters. Officers will use this method to forecast criminal cases and take appropriate measures to apprehend offenders.

KEYWORDS —*Crime, Data mining, Clustering, K-Means, Google Map, Web application.*

INTRODUCTION

Today the crime rate is rising day by day and all the government's efforts around the world are directed at raising the crime rate. Using data mining, which is a great field of application with high data set volume, it has proven that it helps with its techniques to take advantage with historical data. This would help the police departments by analyzing the data and gaining from it the information. Data mining is composed of various techniques such as clustering, grouping, prediction, etc.

Clustering used to cluster objects (groups) with identical dataset attributes for further study and one of the clustering algorithms is k-means algorithm. Most of the crimes are unsolved and contain empty values so that we are limited in choosing clustering over classification. Clustering is regarded as the easiest and fastest clustering algorithm compared to other algorithms and is the best choice for clustering large data volume due to its lower computational speed. In this project we will create a web application with five features that can be useful for the police and this project will use the crime dataset that belongs to Chicago city in the USA and we will implement the k-means algorithm in NetBeans software, Clustering the crime dataset and giving the results that will show us the crime-prone areas with the crime rate that occurred there, and determining the percentage of a particular crime that occurs in a particular area, as well as visualizing the results of each crime on the Google map, including the type, number and time required for police to take advanced crime reduction strategies.

LITERATURE REVIEW

Jyoti Agarwal et al [1] proposed a system which includes steps for crime analysis starting with extraction of crime patterns and prediction the crime using k-means algorithm and that lead to detection of the crime at the end with the use of Rapid miner tool.

Manish Gupta et. al. [2] study the existing system in use by Indian police and highlighted the basic features of the system, then he introduced a criminal analysis tool which is based on data mining techniques that helps   police department to carry

out the activities efficiently.

Maloti and Santhosh Baboo [3] proposed a crime analysis tool for Indian scenario and proved that it is effective for identifying crime patterns and prediction. To develop the tool, he develops a data cleaning algorithm and he enhanced clustering algorithm (Kmeans, DBScan).

Kadhim Swadi Al-Janabi [4] proposed a framework for crime and criminal data analysis. For data classification, he used decision tree algorithms and simple K-means for data clustering. The result is used for prediction of trends and behavior between crimes and criminals.

Chris Delaney [5] he pointed out about the importance of identifying and make tactical analysis of crime trends and patterns by criminal investigators all over the world.

Priyanka Gera and Rajan Vohra [6] did a study at the use of clustering technique in data mining to analyze the crime patterns which helped to recognize which type of crime is more occurring and showed the distribution of each crime type in every area category.

METHODOLOGY

Data mining techniques can be applied in crime analysis and it help to take advantage of historical data and extract knowledge from it therefor help to take better decisions. A data mining approach such as clustering is used to cluster the data into groups where similar objects are placed together and in this system, clustering helps to group the same crime types together which means crime like "murder" are grouped together and same for "rape" cases and so on. In addition to that, the system determines the low, medium and high areas of crime based on the dataset K-means process consist of the following steps:

1. Choose k number of clusters as initial step.
2. Choose a set of K instances as centers of the clusters.
3. Each instance assigned by the algorithm to the cluster which is closest.
4. The cluster centroids are recalculated either after whole cycle of re-assignment or each instance assignment.
5. The process is iterated.

The algorithm requires specifying k number of clusters in advance. It is unable to handle noisy data and outliers and not suitable to discover clusters with non-convex shapes.
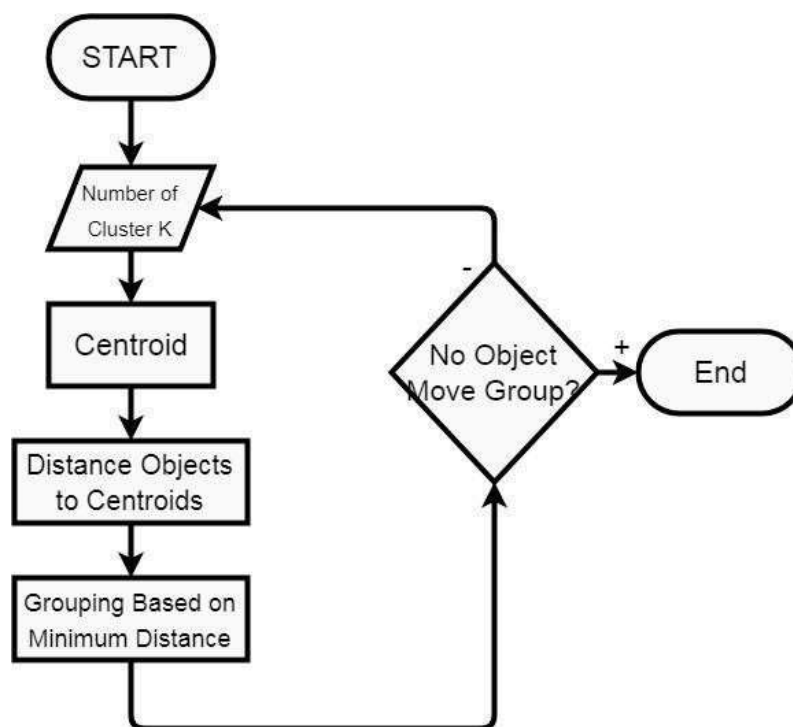
K-means is chosen to be used here because the ease of implementing it using java, plus its simplicity and speed which is very appealing in practice and it is
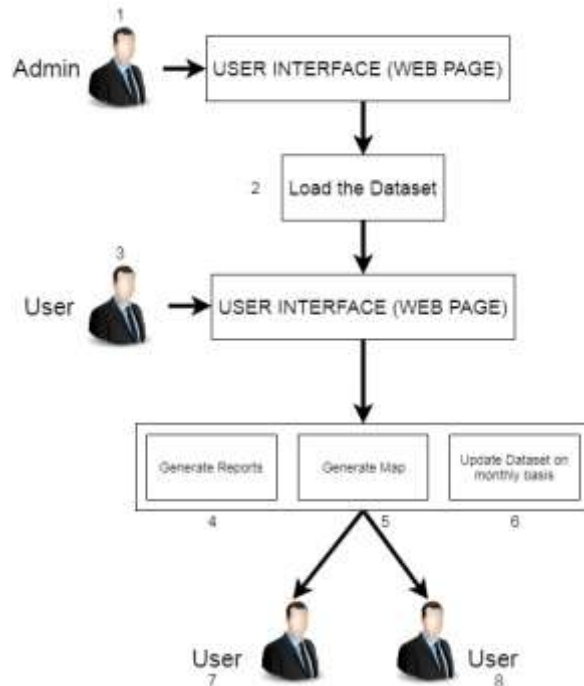

Fig. 2. System Architecture

The dataset imported from Chicago police department included 6.54M rows of committed crimes from 2011 until present and 22 columns such as (id, date, type, location...).

We are going to take this data and imported into NetBeans software and implement the k-means algorithm in java and here we aimed to implement the algorithm in the best way with an optimized code to ensure the efficacy and speed as well as the accuracy in the work.The next part is to choose the number of clusters and it is somehow subjective in general. As we saw, that we are dealing with a huge dataset, we decided to choose 10 clusters for better prediction. Then we can visualize the clusters that are created by k-means on the google map and determine the crime prone areas and each cluster is going to be denoted with a number indicates the crimes included in it. Then it will visualize the places where a specific crime occurs(such as there more theftincidents occurs at tourist destinations and that is the main work which is going to help the police to focus on some areas and take more precautions to prevent any future crimes and hence reduce the crime rate.we can also visualize all the crime locations of the city on the google map for good understanding of the situation. Additional feature is implemented to show the percentage of a specific crime that occur in specific part of the city and an example of that is "rape" incidents are twice in villages than cities. In

suitable for high volume crime dataset and can help to extract useful information.

addition to that, implementing a performance features are necessary to help to monitor the rate of the crime to see if it is increasing or decreasing.
Finally, reports must be generated and can be download that contains the crime rate graphs and map images.

MODULE DESCRIPTION

Proposed system consists of two modules, Admin Module and User Module.

Admin module where his work is to train the dataset and uploaded to make it available to the authorized users. The crime dataset should not contain any empty values and some dataset contain a lot of attributes so here we filter our dataset according to some requirements to ensure the accuracy in the clustering process.

The user module is about the user's work of determining the cluster number which is depends on the method used for measuring similarities and the parameters used for partitioning. Then applying the k-means algorithm on the dataset and starting his job for analyzing the crime patterns and monitoring the crime rate. Then generating the required reports that is going to be demanded by the employees in the police department. So here, many users can share the same work on analyzing the clustering results and trying with different number of clusters to ensure the accuracy and finally they can identify common crime patterns with the help of the system.

CONCLUSION AND FUTURE WORK

This system is primarily used to cluster the dataset and apply the k-means algorithm on the crime dataset for the purpose of crime analysis, and to visualize the clustering results and crime incidents on the map and to prepare crime rate reports. This process is designed to help predict future crime based on the historical datasetMore functionality can be applied to this project in the future as a potential research that will be more useful and provide more knowledge about the crimes and offenders. Various methods may be used in the field of crime analysis by use a particular algorithm to better predict outcomes and to compare the performance and efficiency of the various Machine learning algorithms.

REFERENCES

1. *Jyoti Agarwal, RenukaNagpal, Rajni Sehgal, "Crime Analysis using K-Means Clustering", International Journal of Computer Applications (0975-8887), Vol. 83,No. 04, 2013*
2. *Manish Gupta, B. Chandra and M. P. Gupta, "Crime Data Mining for Indian Police Information System", Computer Society of India, Vol. 40, No.*

1, pp. 388-397, 2008

3. *A. Malathi, Dr. S. Santhosh Baboo, Algorithmic Crime Prediction Model Based on the Analysis of Crime Clusters, Global Journal of Computer Science and Technology Volume 11 Issue 11 Version 1.0 July 2011.*

4. *KadhimB.Swadi al-Janabi . A Proposed Framework for Analyzing Crime DataSet using Decision Tree and Simple K-means Mining Algorithms, Journal of Kufa for Mathematics and Computer, Vol.1, No.3, may2011Y.*

5. *Chris Delaney, Crime Pattern Definitions for Tactical Analysis, International Association of Crime Analysts (IACA), August 2011.*

6. *Priyanka Gera and Rajan Vohra, City Crime Profiling Using Cluster Analysis, International Journal of Computer Science and Information Technologies IJCSIT, Vol. 5 (4), 2014.*

7. *Chicago Police Department (Crime dataset in Chicago city in USA). https://data.cityofchicago.org/Public-Safety/Crimes- 2001-to-present/ijzp*