International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 07 Issue: 03 | Mar 2020 www.irjet.net p-ISSN: 2395-0072

CRIME ANALYSIS AND PREDICTION - BY USING DBSCAN ALGORITHM

Roshan P. Thakur

Department of Computer Science and Engineering, Dr. Babasaheb Ambedkar College of Engineering & Research, Nagpur, Maharashtra, India.

Abstract - *Crime is one of the biggest violations that have been not yet completely solved ever since the evolution of human race. The threat will be reduced if a crime prediction and analysis is concerned on particular areas to find if crime is about to happen or not. The existing system is having trouble about data access speed and it is less efficient so to overcome this problem, we proposed a system in which analysis of crime is done also we use DBSCAN algorithm to find out different cluster of crimes. It shows high accuracy for the given dataset and forms effective cluster. Data mining is an approach that can handle large voluminous datasets and used to predict desired patterns. This aspect will be beneficial for both law enforcement and police organization of our country and to give more correct decision also help in safeguarding an area. This application is useful for police and enforcement of law organizations in order to detect crime and for applying preventive measures.*

***Key Words*: Analysis, Crime; DBSCAN, Existing system; Law enforcement.**

1. INTRODUCTION

Today, time is a concerning factor for sentencing criminals. Many a times criminal is released on bail may yet be a potential threat to the society, even after they have served their sentence. As we all know the rate of crimes is increasing adequately and modern technologies are helping them without knowing such as chatting, videos, news, apps, websites, etc. Criminals cannot be predicted easily so as per the criminals mind not only the persons but also crimes cannot be predicted, so we are analyzing the data we have as a dummy dataset and we are going to conclude these data for future awareness for crime records, it includes city(where it has done), number of crimes happens like these crimes, etc. analyzing the data is difficult because there are no sufficient data is available about a particular crime or there may be inconsistent of data availability. As the crime rates are increasing we be properly analyzed and stored. In this paper, we look at the use of frequent pattern mining with association rule mining to analyze the various crimes done by a criminal and predict the chance of each crime that can again be performed by that criminal. This analysis may help the law enforcement of the country to take a more accurate decision or may help in safeguarding an area if a criminal released on bail is very much likely to perform crime. In this paper we 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. We are going apply these techniques to real crime data. We also use semi supervised learning technique in this paper for knowledge discovery from the crime records and to help increase the predictive accuracy.

1.1 MAJOR CHALLENGES

In the present scenario, following major challenges are encountered:

- Problem of identifying techniques that can accurately and efficiently analyze this growing volumes of crime data.
- To record crime data different methods and structures are used.
- The data available is redundant and there are too many missing values which makes analysis process more difficult.
- Due to complexity of crime issues investigation of the crime takes longer duration.

1.2 AIM AND OBJECTIVE

Aim: Our aim is to develop a user friendly website which can predict regions which have high probability of crime occurrence and can visualize crime prone areas on map.

Objectives: 1.Performing data mining algorithm on available dataset to find required result.2. System can predict areas where there is high possibility and probability for crime occurrence. 3. Visualizing crime prone regions in specific areas.

2. LITERATURE REVIEW

In the study and analysis of criminology data mining can be categorized into two main areas, crime control and crime suppression. In crime control, we use knowledge from the analyzed dataset. And crime suppression used to catch criminal by using his/her

history records. As per the literature survey, crime data is growing very fast and in large amount (running into zota bytes). So we need advanced and efficient techniques for analysis. According to [1], the crime alert areas can be represented graphically using maps ,which indicates the crime alert in respective area. The clustering methods are implemented and their performance is tested based on accuracy. According to [8], we found that the DBSCAN clustering algorithm is more accurate than K-means algorithm with the help of silhouette coefficient. DBSCAN algorithm forms effective clusters.

So, according to survey we are having problem statement, Existing system do not have provision to predict crime prone regions, less efficiency of algorithm causes trouble in analysis process.

3. CLUSTRING TECHNIQUE

DBSCAN is density-based spatial clustering of applications with noise. The DBSCAN algorithm is basically based on clustering points within the distance of epsilon with some initial minimum number of points. [13]It requires epsilon (Eps) as one parameter value and minimum number points as the other parameter. (MinPts).It begins with a random point as its starting point. It then identifies and joins all the nearby points within distance Eps of that particular starting point. A cluster is formed when the number of nearby points joined is greater than or equal to MinPts. If the nearby point is less than the minimum number of points the particular starting point is declared as noise. The start point is then marked as visited. The algorithm repeats the evaluation process for all the neighbors' repeatedly. If the number of neighboring nodes is less than MinPts, the point is marked as noise.

The algorithm is as follows:

- **DBSCAN** is a density based clustering algorithm that works by successively growing a cluster from initial seed points **[1]**.

- If the density in the circle proximity (which has the radius parameter *Eps*) of a point is above or equal a threshold level, denoted by the *MinPts* parameter, the cluster is expanded forward by assigning all the unassigned points in the neighborhood to it.

```
var dbscanner = jDBSCAN()
        .eps(0.075)
        .minPts(1)
        .distance('EUCLIDEAN')
        .data(point_data);
```

- The algorithm then recursively proceeds with the same steps for each of the newly added points to the cluster.

```
// This will return the assignment of each point to a cluster number,
// points which have  -1 as assigned cluster number are noise.
var point_assignment_result = dbscanner();

// (OPTIONAL) If you need the cluster centers for each of the
// identified clusters use this
var cluster_centers = dbscanner.getClusters();
```

- Points that will not be assigned to any cluster by the end of this process are labeled as noise.

4. DATASET USED

The dataset that, we are using was downloaded from kaggle.com. The dataset has details regarding the crime in Boston. The dataset contains attributes such as date, time and day of crime occurrences, crime type, location of crime, etc. For map view dataset contains latitude and longitude to represent crime prone area on map.

Fig -1: Dataset

5. IMPLEMENTATION AND RESULTS

We propose a system which can analyze, classify and predict various crimes, find probability of crime occurrences in a given region. Our system is effective in terms of analysis, speed of crime, classify crime according to their type and show probability of crime occurrences in nearby location.

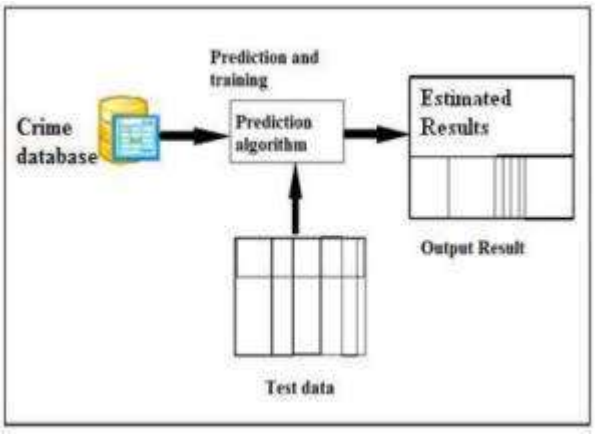

Fig -2: Data processing Steps

Following are steps in doing crime analysis:-

1) DATA COLLECTION:

We use dummy dataset because of limitation in getting crime data records from Law Enforcement department. The collected data is stored into database for further process. Since the collected data is unstructured data we use mango DB. Crime data is an unstructured data since the no of field, content, and size of the document can differ from one document to another the better option is to have a scheme less database.

2) CLASSIFICATION:

For classification we are using an algorithm called Naïve Bayes which is a supervised learning method as well as a statistical method for classification. The algorithm classifies crime based on given training dataset. It is simple, and converges quicker than logistic regression

3) CRIME CLUSTURE:

For clustering data we are using DBSCAN clustering algorithm. The DBSCAN algorithm is basically based on clustering points within the distance of epsilon with some initial minimum number of points. On comparing with K-means the DBSCAN clustering has high accuracy for the given dataset and forms effective clusters.

4) VISUALIZE:

The crime prone areas can be graphically represented using a **GOOGLE-MAP-API** key, also clusters as a predicted regions are also shown in another map view. Visualization helps to get more accuracy in performance.

5.1 MODULES

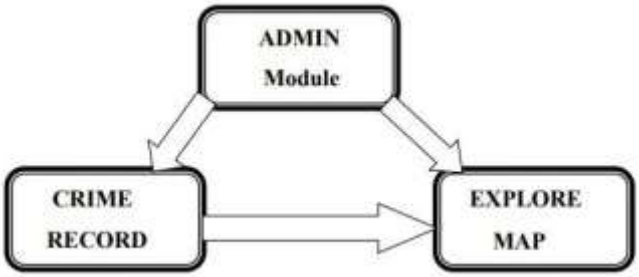

Fig -3: Project Module

1. ADMIN MODULE: Admin will register into the system first to generate User ID and Password. As an admin he/she has authorization to register new crime, update information of crime. Admin can also edit his/her profile and reset password only if he/she is logged in to the system.

2. CRIME RECORD MODULE: This module contains detailed information about crime, Such as criminal name, victim gender, crime location (map), crime type, date and time of crime, for e.g. Crime happens at Night. All the necessary information will be stored in this module and the classification and clustering algorithm will apply on this dataset.

3. EXPLORE MAP: It contains two map view, and one list view. Where, in one map cluster view of crime is shown which helps to find out number of crime register in a particular area and another map view shows predicted region or area of crime. List view where description of crime is given along with date and time registered for that crime. We can use the filters like crime date and time and according to type of crime we can fetch crime from dataset.

5.2. EXPERIMENTAL SETUP

Step 1: Create a new server on the web hosting sites available.
Step 2: Create two databases; one for storing the details of the authorized user and the other for storing details of the crime occurring in a particular location.
Step 3: The data can be added to the database using SQL queries.
Step 4: Create PHP scripts to add and retrieve data. The project is implemented by following steps: Step 1: The

PHP file to retrieve data converts the database in the JSON format.
Step 5: This JSON data is parsed.
Step 6: The location is added by the user by clicking on the map from there latitudes and longitude automatically fetch that is further added to the database.
Step 7: The added locations are marked on the Google map.
Step8:The various crime types used are Robbery, Kidnapping, Murder, Burglary and Rape. Each crime type is denoted using a different color marker.
Step 9: The crime data plotted on the maps is passed to the DBSCAN algorithm.
Step 10: A different colored circle is drawn for different clusters by taking the centroid of the cluster as the center where the color represents the frequency of the crime.

5.3 RESULT

Fig -4.1: Home Page

Fig -4.2: Registration Form for admin

Fig -4.3: Admin Login

Fig -4.4: Map showing crime prone area

Fig -4.5: Add crime record

6. CONCLUSION

We proposed a system in which different patterns and techniques are combined to generate a result, which is used by security agencies and police organizations to easily and economically categorize and analyze crime data to identify patterns and trends.

Our system is effective in terms of analysis speed, identifying common crimes and crime-prone areas for future prediction. As different crimes require different precautions, this can be easily achieved by using this system. It can visualize and give us information about those crimes in a particular region and this system is accessed by common people, security agencies, police organizations, etc.

REFERENCES

[1] Devan M.S (TCS),Crime Analysis and Prediction Using Data Mining Conference Paper , August 2014.

[2] Malathi. A and Dr. S. SanthoshBaboo. Article:an enhanced algorithm to predict a future crime using data mining. International Journal of Computer Applications, 21(1):1–6, May 2011. Published by Foundation of Computer Science.

[3] Anshu Sharma, Raman Kumar, Analysis and Design of an Algorithm Using Data Mining Techniques for Matching and Predicting Crime, IJCST Vol. 4, ISSue 2, AprIl - June 2013.

[4] JyotiAgarwal, RenukaNagpal, RajniSehgal, "Crime Analysis using K-Means Clustering", International Journal of Computer Applications(0975-8887), Vol. 83, No. 04, 2013

[5] AnishaAgrwal, DhanashreeChoghule, ArpitaAgrwal, Application of Crime Data Minning Using Data Minning Method, International Journal of Advanced Computational Engineering and Networking, ISSN: 2320-2106, Volume-4, Issue-5, May.-2016

[6] Veenet Jain, Yogesh Sharma, VaibhavArora, AyushBhatiya, Crime Prediction using Kmeans Algorithm, Global Research and Development Journal for Engineering | Volume 2 | Issue 5 | April 2017.

[7] 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.

[8] Raghavendhar T.V, Joslin Joshy, Mahaalakshmi R, Ashutosh Soni M, Crime Prediction and Analysis using Clustering Approaches and Regression Methods | IJCAT - International Journal of Computing and Technology| Volume 5| Issue 4, April 2018