

# CRIME RATE PREDICTION USING DATA CLUSTERING ALGORITHMS

Omkar Vaidya<sup>1</sup>, Sayak Mitra<sup>2</sup>, Raj Kumbhar<sup>3</sup>, Suraj Chavan<sup>4</sup>, Mrs. Rohini Patil<sup>5</sup>

<sup>1,2,3,4,5</sup>Department of Computer Engineering, Terna Engineering College, Mumbai University, India

\*\*\*

**Abstract** -Crime analysis and prediction is a systematic approach for analyzing and identifying different patterns, relationships and trends in crime. The regions with high probability of occurrence of crime is predicted by the system. The system developed will help to speed up the process of solving crimes for the law enforcement agencies. Data clustering algorithms and data mining techniques are used to extract unknown but useful information from unstructured data. The current data from the police is used which after using different clustering and prediction algorithms provides an insight that will help to predict the likelihood of incidents, track criminal activities and help the law enforcement authorities to deploy resources and also solve crime cases at a faster rate.

**Key Words:** Crime-patterns, data clustering, data mining, k-means, fuzzy c.

## 1. INTRODUCTION

In today's world, crime rate is increasing at a higher rate. It is also difficult to predict the happenings of crime due to several different factors. Despite the use of different modern technologies available, the crime rate has not curbed to a considerable extent. It is difficult to predict like which criminal would do the crime, but the place and the probability of the occurrence of the type of crime can very well be predicted.

The process of solving crimes has been the prerogative of the criminal justice and law enforcement specialists. The increase in the use of computerized systems to track crimes, the law enforcement officers are helped by computer data analysts to speed up the process of solving crimes. In addition to the methods used today, we develop a data mining paradigm that helps to solve crimes at a more faster rate. In order to identify the crime patterns, we use clustering algorithms to get us various clustering models.

For the purposes of our modeling, we will not need to get into the depths of criminal justice but will confine ourselves to the main kinds of crimes. Cluster (of crime) has a special meaning and refers to a geographical area of crime. These clusters of crime can be plotted geo-spatially on a map with the exact location co-ordinates. This will also help to visually represent densely populated areas of crime, in other words "hotspots" of crime. However, when we talk of clustering from a data-mining standpoint, we refer to similar kinds of crime in the given geography of interest. Such clusters are useful in identifying a crime pattern.

The accuracy of the results obtained by prediction will help to a great extent in the process to curb crime and take measures accordingly. So different types of crime records need to be evaluated to develop an efficient crime analysis tool.

The crime data available is present in different forms and is in structured form (written criminal record) as well as in unstructured form (image and audio data from surveillance and different forms and communication). The collected data will be brought together and the volume of the data will contribute to the understanding of the events happened in past and based on the patterns what is more likely to happen in future. With such kind of prediction available, it can help the law enforcement agencies and take actions by understanding the following:

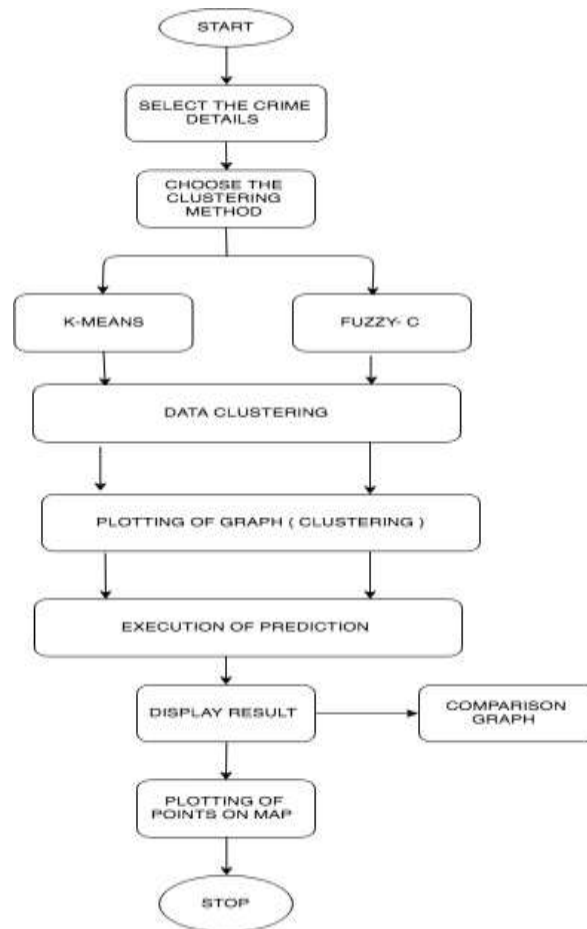
- Identify areas typically frequented by violent criminals.
- Match trends in regional or national gang activity with local incidents.
- Profile crimes to identify similarities and match the crimes to known offenders.
- Identify the conditions most likely to trigger violent crime, and predict when and where these crimes may occur in the future.

## 3. PROPOSED SYSTEM:

The steps for the work flow of the proposed system are as follows:

- a) Select the crime details
- b) Choose the clustering method
- c) K-Means
- d) Fuzzy-C
- e) Data clustering
- f) Plotting of Graph

- g) Execution of prediction
- h) Display Result
- i) Plotting of map



**a) Select the crime details:**

The first step is to select the crime details from the list of attributes available. The dataset is available to use. From the dataset select the desired attributes as such.

**b) Choose the clustering method**

The next step is to select the clustering method. Two options are available

- 1. K-means
- 2. Fuzzy-c

Both the clustering algorithms are well efficient. So it's totally on the user to select which one to be used for clustering of data.

**c) K-means**

This algorithm uses k clusters and centroids from the dataset. It then calculates the distance between the clusters and accordingly places the points in the respective data clusters based on the closest distance.

**d) Fuzzy C**

This algorithm assigns membership to each data point corresponding to each cluster center on the basis of distance between the cluster center and the data point. More the data is near to the cluster center more is its membership towards the particular cluster center.

**e) Data Clustering**

After selecting the criminal records and choosing the clustering method viz. k-means and fuzzy-c by the user, the actual clustering algorithm is run and the clustering of data takes place.

#### f) Plotting of graph

Based on the clustering algorithms, the result is generated and is plotted and is shown in pictorial way in the form of graph.

#### g) Execution of Prediction

The next step after clustering is prediction. This is a very important process in data mining. The crime is predicted by using the Naive Bayes prediction algorithm.

#### h) Display Result

After the prediction step, the data will be displayed on the screen and will show all the details related to the crimes.

#### i) Plotting of data points on map

After getting the clusters from clustering algorithm. The data points are geospatially plotted. The data points are plotted on the map which also shows crime details at that particular location.

### 4. METHODOLOGY

#### 4.1 K-MEANS

K-means is one of the simplest unsupervised learning algorithms that solve the well known clustering problem. The procedure follows a simple and easy way to classify a given data set through a certain number of clusters (assume  $k$  clusters) fixed a priori. The main idea is to define  $k$  centers, one for each cluster. These centers should be placed in a cunning way because of different location causes different result. So, the better choice is to place them as much as possible far away from each other. The next step is to take each point belonging to a given data set and associate it to the nearest center. When no point is pending, the first step is completed and an early group age is done. At this point we need to re-calculate  $k$  new centroids as barycenter of the clusters resulting from the previous step. After we have these  $k$  new centroids, a new binding has to be done between the same data set points and the nearest new center.

Let  $X = \{x_1, x_2, x_3, \dots, x_n\}$  be the set of data points and  $V = \{v_1, v_2, \dots, v_c\}$  be the set of centers.

- 1) Randomly select 'c' cluster centers.
- 2) Calculate the distance between each data point and cluster centers.
- 3) Assign the data point to the cluster center whose distance from the cluster center is minimum of all the cluster centers..
- 4) Recalculate the new cluster center using: where, 'ci' represents the number of data points in ith cluster.
- 5) Recalculate the distance between each data point and new obtained cluster centers.
- 6) If no data point was reassigned then stop, otherwise repeat from step 3).

#### 4.2 FUZZY - C

This algorithm works by assigning membership to each data point corresponding to each cluster center on the basis of distance between the cluster center and the data point. More the data is near to the cluster center more is its membership towards the particular cluster center. Clearly, summation of membership of each data point should be equal to one. After each iteration membership and cluster centers are updated

Let  $X = \{x_1, x_2, x_3 \dots, x_n\}$  be the set of data points and  $V = \{v_1, v_2, v_3 \dots, v_c\}$  be the set of centers.

- 1) Randomly select 'c' cluster centers.
- 2) Calculate the fuzzy membership ' $\mu_{ij}$ ' using:
- 3) Compute the fuzzy centers ' $v_j$ ' using:

4) Repeat step 2) and 3) until the minimum 'J' value is achieved or  $||U(k+1) - U(k)|| < \beta$ .

where, 'k' is the iteration step.

' $\beta$ ' is the termination criterion between [0, 1].

'U = ( $\mu_{ij}$ ) $n \times c$ ' is the fuzzy membership matrix.

'J' is the objective function.

### 4.3 BENEFITS

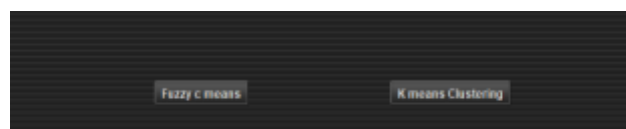
The system helps to predict the criminal and the crime location accurately. The data points are plotted geospatially on the map. The system also helps to calculate and plotted geospatially on the map. The system also helps to calculate and predict the crimes faster. The graphical plotting helps to understand the system. It specifies the estimated time of processing of these data using prediction algorithms.

## 5. IMPLEMENTATION

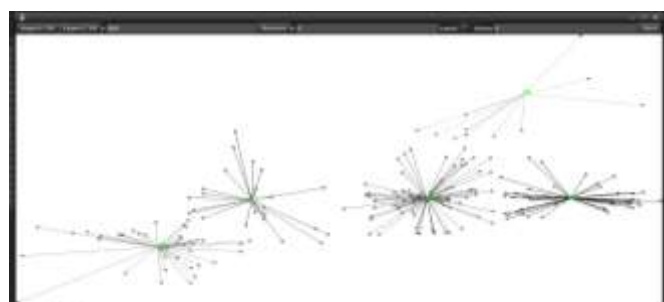
### 1. Selection of crime details



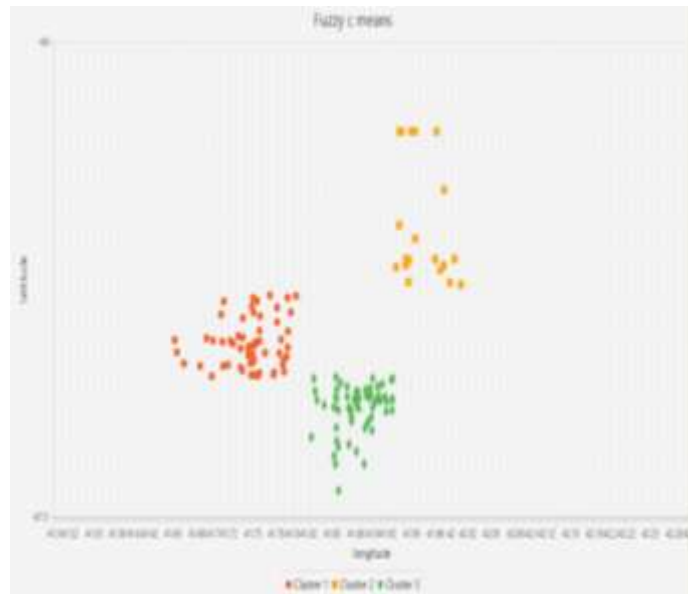
### 2. Choosing the Clustering method



### 3. K-Means Plotting



#### 4. Fuzzy-C Means Plotting



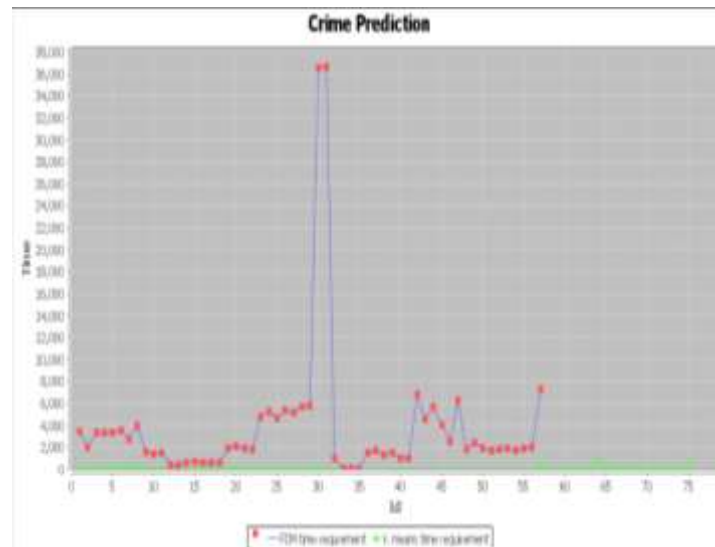
#### 5. Display of data

ID	Case No.	Date	Block	Block	Block	Primary Ty.	Description	Location D.	Arrest	Domestic
1	157075	03/24/14	882991 S.	02500 W 3.	1691	OTHER OFF.	OTHER CRIM.	STREET	FALSE	FALSE
2	157076	03/22/14	882991 S.	02500 W 3.	1320	CRIMINAL	TO VEHICLE	STREET	FALSE	FALSE
3	157077	03/24/14	882991 S.	01300 W 5.	640	THEFT	FINANCIAL	RESIDENCE	FALSE	FALSE
4	157078	03/24/14	882991 S.	02500 W 3.	628	THEFT	DRY AND	SMALL RE.	FALSE	FALSE
5	157082	03/24/14	882991 S.	01800 N.W.	690	MOTOR VE.	AUTOMOB.	STREET	FALSE	FALSE
6	157093	03/25/14	882991 S.	01800 W 5.	690	MOTOR VE.	AUTOMOB.	STREET	FALSE	FALSE
7	157095	03/24/14	882991 S.	01800 N.W.	690	BURGLARY	FORCIBLE	RESIDENCE	FALSE	FALSE
8	157098	03/24/14	882991 S.	05100 W 6.	690	BURGLARY	FORCIBLE	RESIDENCE	FALSE	FALSE
9	157097	03/25/14	882991 S.	05000 N.W.	1310	CRIMINAL	TO PROP.	APARTMENT	FALSE	TRUE
10	157099	03/22/14	882991 S.	04000 W 2.	2825	OTHER OFF.	HANDSOM.	RESIDENCE	FALSE	TRUE
11	157099	03/22/14	882991 S.	02300 W 5.	690	MOTOR VE.	AUTOMOB.	STREET	FALSE	FALSE
12	157095	03/24/14	0282991 S.	00000 E.O.	690	THEFT	OVER \$500	DEPARTME.	FALSE	FALSE
13	157099	03/22/14	882991 S.	00000 E.O.	690	THEFT	OVER \$500	DEPARTME.	FALSE	FALSE
14	157093	02/26/14	882991 S.	08900 S.D.	480	BATTERY	BW/BL	RESIDENCE	FALSE	TRUE
15	157094	03/25/14	882991 S.	08000 S.C.	690	THEFT	OVER \$500	STREET	FALSE	FALSE
16	157095	03/25/14	882991 S.	06400 S.L.	2820	OTHER OFF.	TELEPHON.	RESIDENCE	FALSE	TRUE
17	157099	03/25/14	882991 S.	04000 W 5.	690	MOTOR VE.	AUTOMOB.	STREET	FALSE	FALSE
18	157097	03/22/14	882991 S.	02400 W 1.	628	BURGLARY	INLAWY.	RESIDENCE	FALSE	FALSE

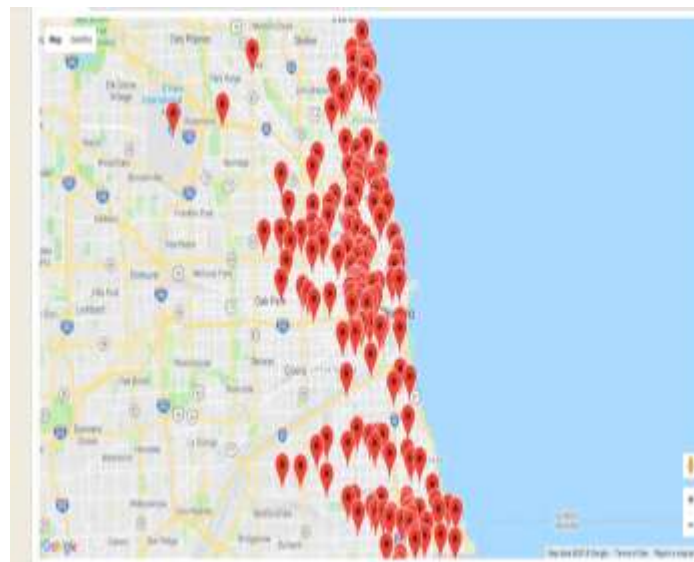
#### 6. Execution of Prediction and Obtaining Result



7. Plotting of Efficiency Graph



8. Geospatially Plotting of Result



6. CONCLUSION

In this paper, we used different data clustering algorithms to cluster data. The clustered crime data is then used to make a prediction and is made to handle changing crime patterns. Crime patterns cannot be static since patterns change over time, so different clustering techniques like K-Means; Fuzzy C etc. are used to handle the changing crime patterns. More in depth inputs will provide better and accurate results.

7. REFERENCES

[1] Shiju Sathyadevan, Devan M.S and Surya Gangadharan. S: Crime Analysis and Prediction Using Data Mining, 2014 First International Conference on Networks & Soft Computing.

[2] Shyam Varan Nath : Crime Pattern Detection Using Data Mining.

[3] Jyoti Agarwal, Renuka Nagpal and Rajni Sehgal : Crime Analysis using K-Means Clustering, International Journal of Computer Applications (0975 – 8887) Volume 83 – No4, December 2013.

[4] Navjot Kaur : Data Mining Techniques used in Crime Analysis:- A Review, International Research Journal of Engineering and Technology (IRJET) Volume: 03 Issue: 08 Aug-2016.

- [5] Adeyiga J.A: A Review of Different Clustering Techniques in Criminal Profiling, International Journal of Advanced Research in Computer Science and Software Engineering -Volume 6, Issue 4, April 2016.
- [6] C.M. FUNG KE WANG AND MARTIN ESTER: - Hierarchical Clustering Algorithms.
- [7] ANDREW MOORE: - K MEANS AND Hierarchical Clustering Algorithms.
- [8] BERND FIRTZKE :- Self Organizing method with constant neighbourhood range and adaptation strength Volume 2, Issue 5, September 1995.
- [9] THOMAS VILLMANN,RALF DER:-Transactions of neural network, Volume 08, No 2, March 1997.