

## Crime Rate Prediction Using K-Means

Snehal Dhaktode<sup>1</sup>, MiralDoshi<sup>1</sup>, Neeraj Vernekar<sup>1</sup>, Ditixa Vyas<sup>1</sup>

<sup>1</sup>(Computer, Atharva College of Engineering, University of Mumbai, India)

---

**Abstract:** Crime is an alarming aspect of our society, and its prevention is a vital task. Crime analysis is a well-organised way of detecting and examining patterns and trends in crime. It is of utmost importance to study reasons, consider different factors and determine the relationship among various crimes occurring and discover the best suitable methods to control crime. The primary objective of this project is to distinguish various crimes using clustering techniques based on the occurrences and regularity. Data mining is used for analysis, investigation and check patterns in crimes. In this project, a clustering approach is used to analyse the crime data; the stored data is clustered using the K-Means algorithm. After the classification and clustering, we can predict a crime based on its historical information. This proposed system can indicate regions which have a high probability of crime rate and distinguish areas which have a higher crime rate.

---

### I. Introduction

Due to the increasing crime rate, it is of utmost importance to have protection against criminal activities. Data mining algorithm can be used to cluster various types of crimes introduced in society, which can be further used to analyse and distinguish crime. This application gives an overview of various crimes which helps dedicated authorities to help reduce and prevent crime up to, certain extent. This project uses K means clustering algorithm to group data based on various factors like time, place, age, type of crime committed of accused. Crime data is stored in the database to perform the analysis. Data mining algorithm will extract information and patterns from the database. We achieve clustering by places where crime has occurred, accused involved in the crime and the time of crime taking place. Admin will enter crime details into the system required for prediction. Admin can view historical criminal data. Crime incident detection depends mainly on the historical crime record.

### II. Procedure

#### I.1. Crime analysis

Crime analysis is an analytic process which provides relevant information about various crime patterns. This project forms clusters using K means algorithm which plays a vital role in analysing and predicting crimes. It clusters various types of crimes by accessing the stored data. It also helps in identifying relevant crime patterns, hidden links, link prediction and statistical analysis of crime data. The prime goal of this study is to examine clustered analysis as an innovative way to detect and help reduce crime and safeguard society. It is vital to analyse crime due to following reasons:

- To inform law enforcers about specific crime trends promptly
- Analyse crime from information existing injustice system and public domain.
- Predict crimes based on existing data and anticipation of using different data mining techniques.

#### 1.2. Data Collection and Classification

Police records contain a large amount of crime data. This data is made accessible by the National Crime Bureau of Records. The data we used is recorded throughout the country over the years. The genuine cases are not filtered and therefore may contain some incorrect values. Processing has to be done for accurate information. According to their characteristics, the data set is categorised into different groups. Grouping is majorly done based on state or city where various types of crimes occur. K-mean algorithm is used to cluster data.

#### 1.3. Potential and Advantages

The crime rate is continually increasing. It cannot be predicted as it is not systematic or random. Even modern technologies work in favour of criminals. According to Crime Records Bureau crimes like burglary, arson has been decreased while crimes like murder, sex abuse, scam, rape have risen. Even though we cannot predict who all may be the victims of crime but can predict the places that have a probability for its occurrence. This proposed system will predict the crime and their patterns of occurrences so that security can be provided for those areas.

### III. Review of Literature

It is only within the last few decades that the technology made spatial data mining a practical solution for broad audiences of Law enforcement officials which is affordable and available. Even today all our data is available in the form of paper-based format. We can digitise this great data record for creating a criminal record database. So the primary challenge in front of us is developing a better, efficient crime pattern detection tool to identify crime patterns effectively.

In countries like England, Cambridge Police Department have done a similar one named Series Finder for finding the patterns in a burglary. For achieving this, they used the modus operandi of an offender, and they extracted some crime patterns which were followed by the offender. The predicted result showed more than 80% accuracy. We are applying a similar concept in this study is to examine the use of clustering technology.

The main challenges we are facing are:

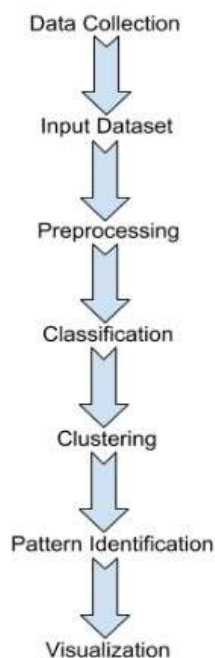
- Increase in crime information that has to be stored and analysed.
- Analysis of data is difficult since data is incomplete and inconsistent.
- Limitation in getting crime data records from the Law Enforcement department.

The precision of the program depends on the correctness of the training set. Finding the patterns and trends in crime is a challenging factor. To identify a pattern, crime analysts takes much time, scanning through data to find whether a particular crime fits into a known pattern. If it does not fit into an existing pattern, then the data must be classified as a new pattern. After detecting a pattern, it can be used to predict, anticipate and prevent crime before this clustering algorithms have been used for crime analysis. For instance, one site it is revealed that suspect has black hair and from next site/witness, it is revealed that suspect is youth and from third one reveals that the offender has a tattoo on his left arm. Through the offender details, we can obtain a basic picture of various crime incidents. These days most of the work is manually done with the assistance of different reports which detectives obtain from data analysts and old crime logs.

#### 3.1 Proposed System Architecture

We would be using an open source server-side language such as PHP for the processing part and MySQL database for storage. We would serve the results to the users with front end languages.

1. First, we take crime dataset.
2. Crime recognition.
3. Filter dataset according to admin and user requirements and create a new dataset which has attribute according to analysis to be done
4. Perform k means clustering on resultant dataset formed
5. From result plot data between crimes and get required cluster
6. Analysis can be done on cluster formed.



**Fig.1** System Flow Diagram

**Input Dataset:** In this step, we are gathering data from registered criminal records and feed it into the system. The user also has the option to register and enter the data himself which can be further verified by the administrator. The collected unstructured data is then stored into the database for future use.

### Pre-processing

Data pre-processing is a data mining technique that involves transforming raw data into an understandable format. Real-world data is often incomplete, inconsistent, and lacking in certain behaviours or trends, and is likely to contain many errors. In this step, we clean the data and make it ready for our clustering algorithm so that it is easy to make predictions and reduce any errors.

### Classification

Classification is a step-in data mining that allocates target categories or classes to the items in the collection. We can correctly predict the destination class for each case in the data. In this case, a classification model is used to group crimes based on various parameters.

For example, Crimes can be classified based on severity - Low Risk, High risk.

### Clustering

It is used to determine association rules which highlight general trends in the database. In this step, we identify the crime pattern for a particular place and correspond to each location we take the attributes and predict some pattern to that particular cluster. When a new case comes, it follows the same pattern and says that this area has a chance for crime occurrence.

### Crime prediction

It is simple to understand and interpret and it works well with large datasets.

## IV. Setup and Results

### 4.1. Approach Used

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

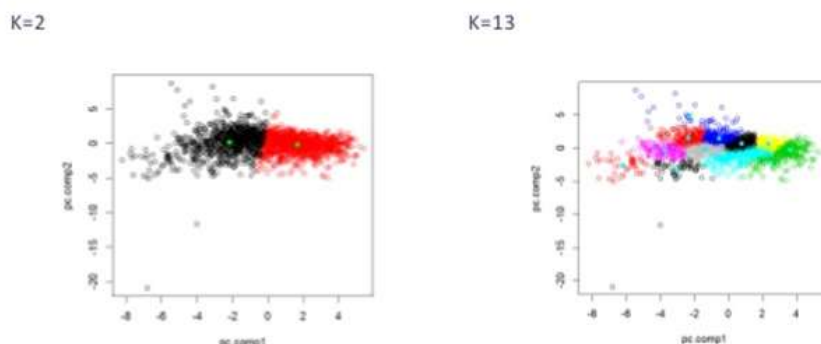
### 4.2. Process

The initial step is the choice of a set of  $K$  instances as centres of the clusters. Next, the algorithm considers each instance and assigns it to the closest cluster. The cluster is recalculated either after the whole cycle of re-assignment or each instance assignment. This process is iterated.

### 4.3. Cluster

K means cluster analysis involves tracking crime rate changes throughout the years, which is further used in data mining to project those changes into the future. Here we consider homicide crime and plot it with year and analysis variation in a graph on cluster formed.

## K-MEAN Clustering



**Fig 2** (a) Non Classified Clusters (b) Classified Clusters

#### 4.4. Result

The result of this research will be to analyse, correlate and predict the crimes from huge data set available. Results will be in the form of correlation between various crime and location of a crime, i.e. state/city. Crime can also be correlated by age group, a location of crime & type of crime occasion of the crimes and frequently occurring crimes. Prediction of the crime will be displayed using various diagrams pie charts and tables.

#### 4.5 Scope

In this proposed system the inputting some test data into the model which shows better results. This system takes factors/attributes and gives frequent patterns of that place. The pattern is used for building a model for decision tree. Corresponding to each place we build a model by training on these various patterns. Crime patterns cannot be static since patterns change over time. By training means we are teaching the system based on some particular inputs. So the system automatically learns the changing patterns in crime by examining the crime patterns. Also, crime factors change over time. By sifting through the crime data, we have to identify new factors that lead to crime though full accuracy cannot be achieved. For getting better results in prediction, we have to find more crime attributes of places instead of fixing specific attributes. Thus, the proposed system using certain attributes but we are planning to include more factors to improve accuracy

### V. Visualization

Security is an aspect which must be given higher priority by all political and government organisations worldwide, aiming to reduce crime incidences. As data mining is the appropriate field to apply on more massive datasets, the knowledge gained from data mining approaches will be highly useful and will support the police force. So, In this paper crime analysis is done by performing k-means clustering on crime dataset.

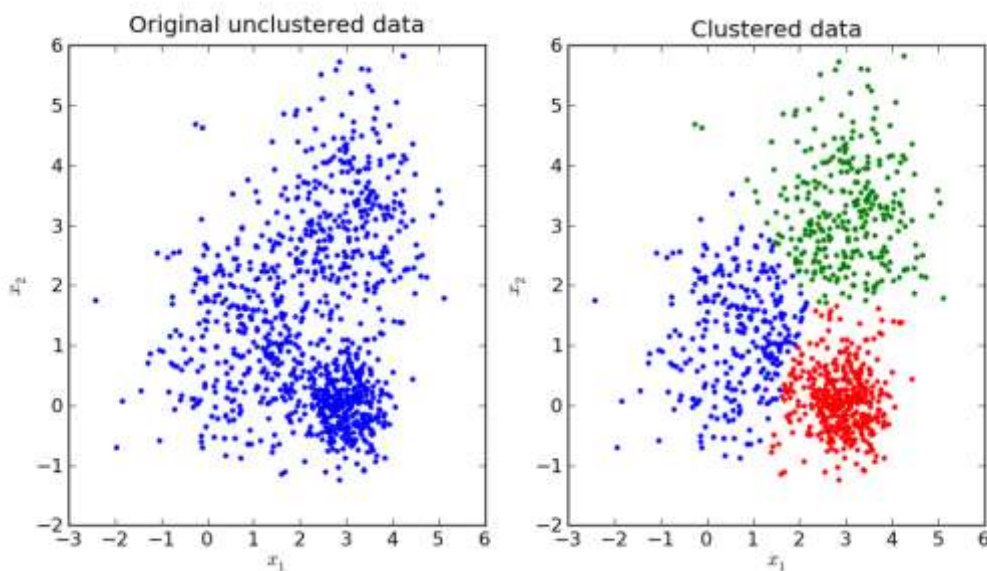


Fig 3 Clustered and Unclustered Data

### VI. Implementation Plan

As we need to perform unsupervised learning on unlabelled data, we use K-Means clustering; The variable K represents the number of groups. Based on the features provided in the data, the algorithm works iteratively to assign each data point to one of the K groups. Data points are clustered based on feature similarity. The results of the K-means

The clustering algorithm is

1. The centroids of the K clusters, which can be used to label new data
2. Labels for the training data (each data point is assigned to a single cluster)

Rather than defining groups before looking at the data, clustering allows you to find and analyse the groups that have formed organically.

**Table 1 - Database Example Table**

ID	Integer
Crime	Char
Location	Char
Time	Datetime
Occasion (if any)	Varchar
Purpose	Varchar
Accused	Char

## **VII. Future Scope**

As of now, the project relies on manual input from a human (a police officer) in order to enter details in the database. If we can make this a centralised system and connect it to all the police stations countrywide and make FIR reporting digital, then it would be quite easier to predict crimes in that particular location and recognise patterns in them. It would also encourage citizens to track their E-FIR online. We can also avoid corruption as the government can keep a track on the number of cases registered and their solvability rate which can help them utilise their resources better.

## **VIII. Conclusion**

In this system, we get to classify and cluster to improve the accuracy of location and pattern-based crimes. This software predicts frequently occurring crimes and its locations, especially for particular state, region and occasions. Also, time is an essential factor in the occurrence of crimes, and we shall predict the time as well in this application.

## **Acknowledgement**

We owe sincere thanks to our college Atharva College Of Engineering for giving us a platform to prepare a project on the topic “CRIME RATE PREDICTION USING K - MEANS” and would like to thank our Principal Dr. Shrikant Kallurkar for instigating within us the need for this research and giving us the opportunities and time to conduct and present a research on the topic.

We are sincerely grateful to Prof. Mahendra Patil, Head of Computer Engineering Department, and our project coordinators Prof. Deepali Maste, and Prof. Mamta Meena during our research on “CRIME RATE PREDICTION USING K - MEANS.”

which would have seemed difficult without their motivation, constant support and valuable suggestions.

Moreover, the completion of this research would have been impossible without the cooperation, suggestions and help of our friends and family.

## **References**

- [1]. Detection Analysis & Prediction Meet Meet Timbadia, Ajit Yadav Information Technology, University Of Mumbai, Shree L.R.Tiwari College Of Engineering, Thane, India, Year – 2017.
- [2]. Z-crime: A Data Mining Tool for The Detection of Suspicious Criminal Activity Based on Decision Tree Mugdha Sharma Information Technology, University of Mumbai, Shree L.R Tiwari College of Engineering, Thane, India, Year – 2015.
- [3]. A Survey of Data Mining Techniques for Analyzing Crime Patterns Ub on Than sataomwatan. Second Asian conference on defence technology. Year -2017.
- [4]. Crime Prediction and Forecasting in Tamilnadu Using Clustering Approaches S Sivaranjani, S Sivakumari, M Aasha. Dept. Of information technology. University of Mumbai L.R. Tiwari College Year- 2017.
- [5]. 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.