

# Analysing Crime Data Using K-Means Clustering

Paras Madan

*Paras Madan, Maharaja Agrasen Institute of Technology, New Delhi*

Submitted: 10-06-2021

Revised: 20-06-2021

Accepted: 23-06-2021

**ABSTRACT** - In today's world the increasing crime rate is one of the most alarming dangers to the society. Government and political bodies are also serious about the issue and are willing to reduce the crime rates. We can analyze crimes in different regions and use k-means clustering algorithm to categorize these regions based on crime rates which will support both the citizens and the police force.

**Key Words:** Crime, K-Means, Clustering, Crime Analysis, Safety, Data

## I. INTRODUCTION

In today's world, due to the advancements in technology and easily available knowledge online on the internet criminals are becoming technologically advanced in committing any crime. Therefore many challenges are faced by the police, intelligence and law enforcement agencies. One such major difficulty is the amount of crime data that is available which is predominantly due to the high crime rates i.e. difficulty in analyzing large amount of data that is involved in crime activities therefore the government bodies and law agencies need to devise and employ new techniques to catch the criminals thereby reducing the crime rate. In such situations we can take help of machine learning. To perform crime analysis appropriate approach need to be chosen and as clustering is an approach of grouping together similar data and dividing the dissimilar data this is an approach which can be used for analyzing the crime data. These groups are known as clusters and different techniques of clustering are K-means clustering, Mean-Shift clustering, Density-based spatial clustering of applications with noise, Expectation-maximization clustering using gaussian mixture models, Agglomerative hierarchical clustering, etc. In this paper the algorithm of K-means clustering has been used on the scraped crime data using automation so that it useful for police, public and crime cells to reduce further happenings of similar crimes and providing information to reduce the same. Also for crime analysis the dataset has been created by scraping the Times of India news website from 2015-2021.

### 1.1 Crime Analysis

Crime analysis is an important part of the research. Crime analysis refers to extracting the meaningful information which can applied to use or effect from the already existing high volume datasets. It can provide information related to crime patterns and also various trends in crime.

Reasons to analyze crime data:

1. It helps government agencies and law enforcers to develop strategies to counter the increasing crime rates.
2. It can help commuters and general public to travel safely to their destinations.

It is to be noted that the crime patterns are also changing and it essential to develop newer advanced techniques to keep up with the changing 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

## II. APPROACH USED

### 4.1.1 k-means clustering algorithm

K-means clustering is one of the method by which we can cluster our data into k clusters.

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 values which will be the center of the clusters.
3. Next, the algorithm considers each instance and assigns it to the cluster which is closest.

K means algorithm complexity is  $O(tkn)$ , where n is instances, c is clusters, and t is. 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.

### 2.1 Dataset Used

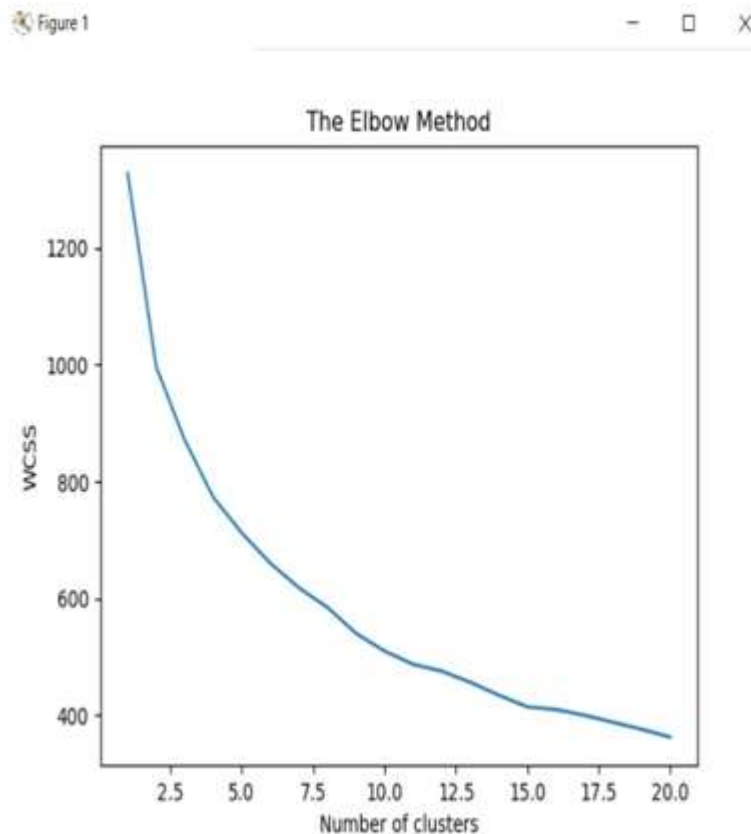
The data was collected using web scraping from the Times of India newspaper.

```

hm_poi,murder,rape,gangrape,robbery,theft,assault_murders,sexual_harassement,totarea,totalcr
CHITRANJAN PARK,2,6,1,35,442,19,7,2659329.537,512,77.2492,28.53632,192.5297308,2.659329537
DABRI,8,28,0,79,240,26,16,3401013.428,397,77.086,28.61268,116.7299125,3.401013428
MALVIYA NAGAR,3,28,1,33,694,63,15,1379853.572,837,77.20418,28.52989,606.5861022,1.379853572
CHANDNI MAHAL,1,8,1,23,529,19,7,5570696.132,588,77.23608,28.64361,105.5523378,5.570696132
MODEL TOWN,0,4,1,45,393,9,14,2689157.085,466,77.19369,28.70257,173.2885009,2.689157085
ANANDVIHAR,10,14,0,99,457,24,15,4558969.692,619,77.29373,28.65335,135.776292,4.558969692
KASHMERE GATE,0,9,3,17,340,17,12,1627909.974,398,77.22618,28.66645,244.4852641,1.627909974
GOVIND PURI,1,4,3,31,692,5,5,8992712.06,741,77.26532,28.53067,82.40005852,8.99271206
BINDAPUR,4,4,1,64,417,11,8,2568079.5,509,77.06618,28.6091,198.2025068,2.5680795
NEW FRIENDS COLONY,6,7,1,47,609,14,10,4045386.71,694,77.2676,28.56234,171.5534385,4.04538671
SARITA VIHAR,1,12,0,12,211,8,1,7942774.025,245,77.30617,28.52299,30.84564653,7.942774025
TIMARPUR,8,17,1,60,626,23,5,6149745.104,740,77.2242,28.7072,120.3301905,6.149745104
KANJHAWALA,1,5,2,15,307,21,10,35490190.9,361,77.00361,28.72582,10.17182469,35.4901909
ANAND PARBAT,5,20,1,77,374,35,9,1948430.286,521,77.17376,28.66004,267.394735,1.948430286
SAGARPUR,2,7,0,11,180,4,4,1283551.172,208,77.08644,28.59139,162.0504149,1.283551172
PRASHANT VIHAR,7,8,0,41,457,39,14,2076678.927,566,77.12757,28.71793,272.5505578,2.076678927
SOUTH CAMPUS,10,11,1,38,248,8,9,4518856.098,325,77.179,28.57051,71.92085629,4.518856098
ROHINI SOUTH,6,14,0,96,560,24,5,2649542.395,705,77.1159,28.70236,266.0836835,2.649542395
  
```

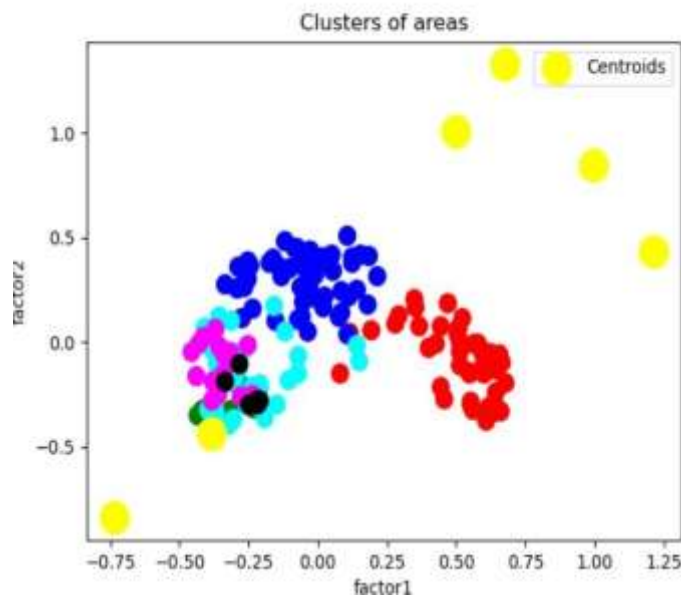
## 2.2 K means cluster analysis

The K-means clustering algorithm was applied on the above dataset and following results were obtained.



Seeing the elbow graph, the value of k was taken to be 6.

Figure 1



### III. METHODOLOGY

After clustering the places using the K-means clustering algorithm. Specific danger indexes from 1 to 6 were assigned to all of the places, with 6 being the most dangerous and 1 being the least. Now using this data, routes were found the google maps between two locations in the city of Delhi. Multiple routes were displayed on the map with different colors. Also, a route summary was shown in which the information of the above found routes were displayed like distance, time of travel, etc. And the route with the least dangerous index was highlighted in green color. To implement this API's like Maps Javascript API, Directions API and Places API were used. Also several modes of travel were provided to the user such as public transport, walking and driving. Some other features were also provided to the users such as tapping an alert button would send an alert message to all the relatives of the victim if he/she is in a dangerous situation. Also a button was also provided to the user by which he/she can see the nearby places in the area using the users location such as hotels, police stations, etc. These all features were combined and an application was made which can help prevent crimes in the city of Delhi. The users would also be able to report the crimes through the portal.

### IV. CONCLUSION

The project focuses on calculating the areas with high and low crime rates depending upon the different types of crimes committed in the different regions of an area. It eventually aims to curb the crime data to ensure the safety of the commuters and general public.

### V. FUTURE SCOPE

In future we can add more parameters to our data to increase the accuracy in our results. We can use this data for different purposes which can help in the safety of the public. It can also be of help in certain areas to the police and crime cells.

### REFERENCES

- [1]. Li, Youguo & Wu, Haiyan. (2012). A Clustering Method Based on K-Means Algorithm. Physics Procedia. 25. 1104-1109. 10.1016/j.phpro.2012.03.206.
- [2]. Jin, Xin & Han, Jiawei. (2011). K-Means Clustering. 10.1007/978-0-387-30164-8\_425.
- [3]. Shi, Na & Liu, Xumin. (2010). Research on k-means Clustering Algorithm: An Improved k-means Clustering Algorithm. 3rd International Symposium on Intelligent Information Technology and Security Informatics, IITSI 2010. 63-67. 10.1109/IITSI.2010.74.
- [4]. Lewis, Chris. (2013). The Truth About Crime Statistics. The Police Journal. 86. 220-234. 10.1350/pojo.2013.86.3.625.
- [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]. Cui, J. & Wang, X.. (2008). Research on Google map algorithm and implementation. Journal of Information and Computational Science. 5. 1191-1200.
- [7]. Dodsworth, Eva & Nicholson, Andrew. (2012). Academic Uses of Google Earth and Google

- Maps in a Library Setting. Information Technology and Libraries. 31. 10.6017/ital.v31i2.1848.
- [8]. Haddaway, Neal. (2015). The Use of Web-scraping Software in Searching for Grey Literature. Grey Journal. 11. 186-190.
- [9]. Zhao, Bo. (2017). Web Scraping. 10.1007/978-3-319-32001-4\_483-1.
- [10]. Parsekar, Shradha & Singh, Mannat & Bhumika, Tumkur & Nair, Sreekumaran. (2015). Road safety in India: A Public Health Concern. Indian Journal of Community Health. 27. 191-6.
- [11]. K. Zakir Hussain, M. Durairaj and G. Rabia Jahani Farzana ,2012 Application of Data Mining Techniques for Analyzing Violent Criminal Behavior by Simulation Model
- [12]. Mohan, Dinesh & Tsimhoni, Omer & Sivak, M. & Flannagan, M.. (2009). Road Safety in India: Challenges and Opportunities. UMTRi-2009-1.