# CRIME ANALYSIS USING K-MEANS ALGORITHM

**Reshma Ratheesh [1], Anita Abraham[2] , Jubiya George[3] and Mrs. Jetty Benjamin[4]**

*123 PG Scholars, Amal Jyothi College of Engineering.*
*4 Assistant Professor, Amal Jyothi College of Engineering.*

**Abstract: Data mining is the process of extracting useful records from the massive quantity of records and changing it into comprehensible form for further use. Clustering is the process of grouping item attributes and features such that data objects in a one group are greater similar than data objects in any other group. But it is a long way now very tough due to the sharply boom within the massive volume of information generated by using number of applications. Fixing those problems is the difficulty of many modern-day research works.**
**Thus, an evaluation on k-mean clustering algorithms is performed.**

**Keywords- Crime analysis, Clustering, K-means**

## I.INTRODUCTION

In these days international protection is a thing which is given higher precedence by all political and government worldwide and aiming to lessen crime incidence. Data mining or knowledge discovery in databases (KDD) is the system of extracting relevant data from huge data in data base. The reason of the data mining approach is to mine information and gain knowledge from a bulky dataset and make over it into a reasonable form for supplementary purpose.

Here,s grouping or clustering of data sets primarily based on the similarity of crime type is performed. This clustering by way of mining the information from the dataset can be beneficial and help police force.

A crime is an offence that merits community condemnation and punishment, usually by way of fine or imprisonment. A crime or offence (or criminal offence) is an act harmful not only to some individual or individuals but also to a community, society or the state ("a public wrong").

Clustering is a procedure of unsupervised learning. . Highly superior clusters have high intra-class similarity and low inter-class similarity. Several algorithms have been designed to perform clustering, each one uses different principle. They are divided into hierarchical, partitioning, density-based, model based algorithms and grid-based.

There exist different types of crime and the grouping of these crimes that occurred in different region, time etc is called clustering. This grouping can be done on different basis. These issues can be solved using the K-means algorithm.

There exist special sorts of crime and the grouping of these crimes that be fell in distinct place, time etc is referred to as clustering. This grouping can be done on specific basis . these problems may be solved using the k-means set of rules.

k-method clustering method is widely used clustering algorithm, which this is utilized in medical and business programs. It is far a technique of cluster analysis that is used to partition N objects into k clusters in this type of way that each item belongs to the cluster with the nearest mean. K-means is faster and clean to enforce. With a massive quantity of variables, k-means may be computationally faster than hierarchical clustering (if k is small). K-means may additionally produce tighter clusters than hierarchical clustering. An instance can change cluster (move to any other cluster) while the centroids are re-computed.

## II.LITERATURE REVIEW

Bharti et al [1] developed a solution to investigate the network of co-offenders in India and predict the possible future network of offenders. The usage of distinctive data mining and graph mining strategies, we will use this method to investigate the community of co-offenders in India and

are expecting the possible destiny community of offenders. Additionally, visible and intuitive crook and intelligence investigation techniques may be developed for crime patterns of metropolis.

Agarwal et al [2] evolved challenge that focuses on crime analysis by implementing clustering algorithm on crime dataset using rapid miner tool and crime analysis by considering crime homicide and plotting it with respect to year . A solution to this problem was developed that crime analysis is done on crime dataset by applying k means clustering algorithm using rapid miner tool. And from the clustered results it is easy to identify crime trend over years and can be used to design precaution methods for future.

Azeez et al [4] proposed a hybrid approach to crime event prediction using Deep learning and its optimization for crime event prediction. As a method to the problem. it's viable to map maximum probable crime activities , its geo spatial, temporal details and institution the type of crime.

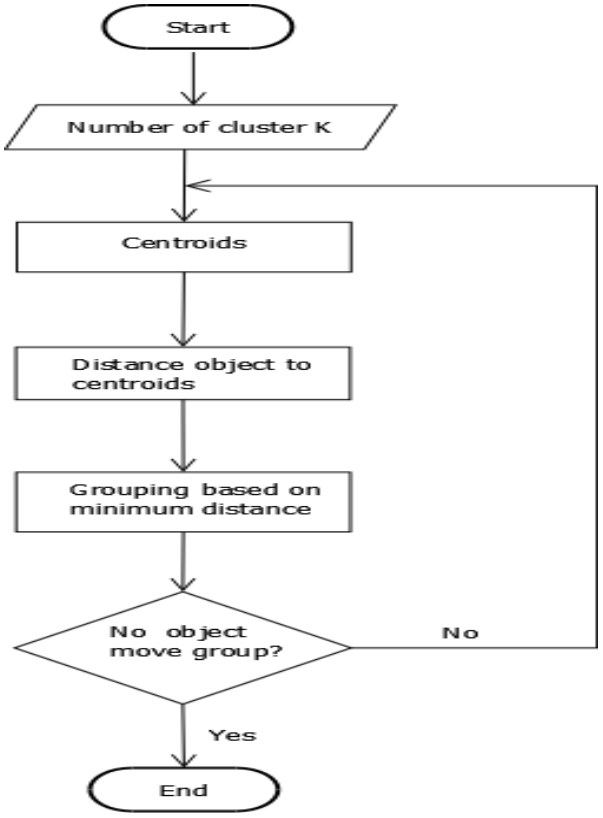
Jain et al [5] evolved a solution to the trouble for a system which can expect areas which have high probability for crime incidence and may visualize crime prone areas. From the obtained the clustered results using the k-means algorithm. It is simple to discover crime susceptible regions and may be used to design precaution techniques for destiny.

## III.IMPLEMENTATION

Input: Number of clusters.

1. To start with, the number of clusters need to be known allow it be k
2. The initial step is to choose a set of K instances as centers of the clusters.
3. Next, the algorithm considers every instance and assigns it to the cluster which is closest.
4. The cluster centroids are recalculated either after whole cycle of re-assignment or each instance assignment.
5. This procedure is iterated. Output: A set of k clusters.

k-means algorithm is a base for all other clustering algorithms to find the mean values.




## IV.METHOD OF IMPLEMENTATION

### Creating data set

```
$n=0;

$sql1=mysqli_query($con,"select*from
`tbl_crime_type`");
while($row1=mysqli_fetch_array($sql1))

{

$c[$n]=$row1['crid'];

$n++;

}
```

### Creating centroid

```
$m=0;

$sql2=mysqli_query($con,"select*from
`tbl_closed_case`");
while($row2=mysqli_fetch_array($sql2))

{

$arr[$m]=$row2['crid'];

$m++;
```

```
}
```

**Calculating the Euclidean distance**

```
for($i=0;$i<$n;$i++)
{
for($j=0;$j<$m;$j++)
{
$diff=$c[$i]-$arr[$j];
$d[$i][$j]=abs($diff);
}
}
```

**Finding the minimum**

```
$k=0; for($i=0;$i<$m;$i++)
{
$min=$d[0][$i];
$a=0; for($j=0;$j<$n;$j++)
{
if($min>$d[$j][$i])
{
$min=$d[$j][$i];
$a=$j;
}
}
$cl[$k]=$a;
$k++;
}
```

**Finding Arithmetic Mean**

```
for($i=0;$i<$n;$i++)
{
$sum=0;
$r=0; for($j=0;$j<$m;$j++)
{
if($cl[$j]==$i)
{
$sum=$sum+$arr[$j];
$r++;
}}
if($r!=0)
{
$s=$sum/$r;
$c[$i]=(int)$s;
}
}
```

**Comparing the centroid**

```
if(array_values($old_c)==array_values($c))
{
break;
}
```

## V.RESULT

This paper work covers an easy and efficient manner for assigning data points to clusters. This work guarantees that the process of grouping takes 0.0023 seconds time complexity while data set is immediately retrieved from the database and takes 0.00000000086 seconds time complexity for clustered data using the algorithm, without sacrificing the accuracy of clusters.

## VI.CONCLUSION

This project focuses on crime clustering based on crime dataset the usage of K-means algorithm. This method is quite simple and efficient for clustering data set. The end result obtained is accurate and can be used for future development and implementation. The time complexity of the work developed could be easily calculated without causing any mistakes in the process using K-means. As a way to improve the accuracy overcoming the restrictions of k-means algorithm, implementation of other algorithms will be greater powerful.

## VII.FUTURE SCOPE

From the outcomes acquired, it is miles clear that the k- mean algorithm has a promising destiny for increase in the effectiveness and efficiency of crook and intelligence analysis. As we have applied clustering technique of data mining for crime evaluation we also can perform other techniques of data mining which includes classification. But it has some restrictions like- it is difficult to predict the number of clusters (k-value) ,the preliminary seeds have a strong effect at the very last outcomes , the order of the data has an impact on the very last outcomes and it is sensitive to scale as rescaling your datasets (normalization or standardization) will completely

change results. Fixing these troubles is the concern of many recent research works. And so as to overcome these limitations and increase the efficiency several other algorithms might be useful .

## REFERENCES

[1] Sushant Bharti , Ashutosh Mishra - "Prediction of Future Possible Offender's Network and Role of Offenders "2015 Fifth International Conference on Advances in Computing and Communications

[2] Jyoti Agarwal, Renuka Nagpal, Rajni Sehgal "Crime Analysis using K-Means Clustering"**, Article** in International Journal of Computer Applications · December 2013

[3] "Top 10 algorithms in data mining" - Received: 9 July 2007 / Revised: 28 September 2007 / Accepted: 8 October 2007 -Published online: 4 December 2007

[4] Jazeem Azeez , D. John Aravindhar - "Hybrid Approachto Crime Prediction using Deep learning"- 2015 InternationalConference on Advances in Computing, Communications and Informatics (ICACCI)

[5] Vineet Jain , Yogesh Sharma , Ayush Bhatia , Vaibhav Arora-Crime Prediction using K-means Algorithm (GRDJE/ Volume 2 / Issue 5 / 035