# IJESRT

## INTERNATIONAL JOURNAL OF ENGINEERING SCIENCES & RESEARCH TECHNOLOGY
## COMPARISON OF CLUSTER BASED TECHNIQUES FOR ZONING OF CRIME INFORMATION

### Ms. Ashwini Gharde[1], Prof. Mrs. Ashwini Yerlekar[2]
1Computer Sci. and Engg, RTMNU/Rajiv Gandhi College of Engg. And Research, Nagpur
2Computer Sci. and Engg, RTMNU/Rajiv Gandhi College of Engg. And Research, Nagpur

## ABSTRACT
Crime is classically "unpredictable". It is not necessarily random, but neither does it take place consistently in space or time. A better theoretical understanding is needed to facilitate practical crime prevention solutions. The crime rate and the activities of criminals have expanded in last few decades due to better communication system and transport.In India, the regional location has a powerful impact on criminal activity. The Crime Info report of National Crime Records Bureau (NCRB), India collects, analyse and publish the crime data. The crime profiling and zoning can be modelled with utilization of data mining.In this paper, we make cluster analysis on the basis of regions like north, south, east, west and central by using k-means cluster algorithm on criminal dataset of India. The cluster input is used to create result analysis graph for each regions along with categories of crime did by male, female and juvenile. Also we are providing comparative analysis of crime rates in different zones of India using KNN and Hybrid classifier algorithm.

**Keywords:** Data mining, crime profiling, clustering, k-means, KNN classifier, Hybrid classifier

## I. INTRODUCTION
We, human beings need to live and work in a safe place which is out of every danger and attack. We want to ensure that there is a concerned authority and systems that protects our lives as well as our properties from potential threats. In fact, one of the main functions of any government is to ensure that law and order for the security of its citizens are put in place. Today, security is one of the major concerns for us and it is given top priority by the National Crime Records Bureau, India. The crime graph of India is constantly increasing. Preferably crime should be prevented, if it cannot be prevented then it should at least be detected and therefore, we are proposing a methodology which helps us to find out the zone with high medium and low risk.

In India NCRB maintains the crime data. In this paper we use cluster technique of k-means algorithm on criminal dataset of India which is collected from NCRB site for the year 2015. The crime dataset is developed by pre-processing the Crime Info database. States are divided into five zones North, East, South, West and Central. Further every zones are classified as Male, Female and Juvenile. Also consider that we have considered only three crimes in Male Female and Juvenile from the vast amount of dataset. Further several categories for crimes has been considered. We use cluster technique of k-means algorithm on criminal Dataset of nature and number of complaints received by police and cases registered under IPC & SLL during 2015. The cluster input is used to create custom India Graph with the cluster zones of states. For every category the Rate parameter is calculated by the formula

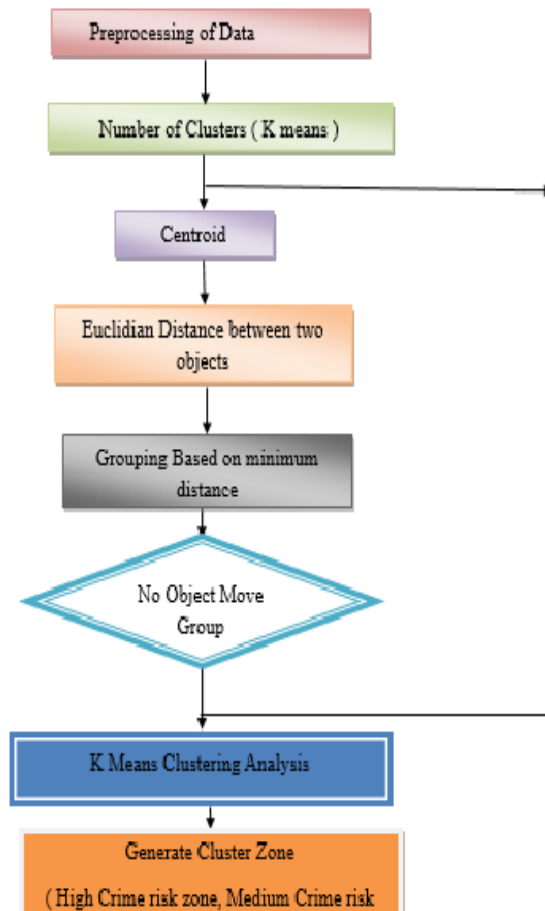Rate= no of crimes/ Total Population

## II.     PROPOSED WORK



*Fig 1: proposed work*

## III.     METHODOLOGY

Once data is collected data pre-processing is done. We removed noisy data and took the rate of crime with respect to the population of male, female and juveniles. Then we store the processed data in the database for the further processing. By using processed data we find Euclidean distance of the rate of crimes according to region.

The formula for the calculating Euclidean distance is as follows.
*dist ((x,y), (a,b))= √(x - a)² + (y - b) ²*

With the help of Euclidean distance we find centroid and this centroid is used for the clustering process by using the k-mean algorithm. K-mean is used for forming the clusters. Steps involved in the k-mean algorithm are:
   1.   Partition of objects into k non-empty subsets.
   2.   Identifying the cluster centroids (mean point) of the current partition.
   3.   Assigning each point to a specific cluster
   4.   Compute the distances from each point and allot points to the cluster where the distance from the centroid is minimum.
   5.   After re-allotting the points, find the centroid of the new cluster formed.

After forming the clusters we use the KNN classifier for finding the high, medium and low crime rates by male, female and juvenile in a particular region i.e. south, north, east, west and central. K-NN classifier uses the nearest neighbour approach for the classification. Steps of KNN classifier:
   •   Calculate "d(x, xi)" i =1, 2... n; where d denotes the Euclidean distance between the points.
   •   Arrange the calculated n Euclidean distances in non-decreasing order.

---

- Let k be a +ve integer, take the first k distances from this sorted list.
- Find those k-points corresponding to these k-distances.
- Let ki denotes the number of points belonging to the ith class among k points i.e. $k \geq 0$
- If ki >kj $\forall$ i $\neq$ j then put x in class i.

And at the end the comparative analysis and performance evaluation of the KNN classifier and hybrid classifier are calculated.

## IV.     CONCLUSIONS

Criminology is a sensitive area where proficient clustering approaches of data mining plays vital role for crime analysts. We focus on criminal analysis by executing k-means clustering algorithm on Crime Info NCRB dataset of India. The clustering method is used to group the states in India according to the criminal data of total, male, female and juvenile number of crimes for the year 2015. The cluster zoning and custom maps generated can help state police and law enforcement department to take additional preventive measures in high and medium crime risk zones to combat against crime and plan advanced investigation strategies. The crime trend and zoning knowledge can also be helpful in cautioning police to increments and reductions in levels of preventive actions. In future, we can likewise perform different methods of data mining on CrimeInfo NCRB India dataset with more criminal attributes to identify crime trend, crime patterns. The knowledge can be used to frame crime control methods and optimal deployment of resource in crime prevention for future

## V.     REFERENCES

[1] A Review of Different Clustering Techniques in Criminal Profiling, Adeyiga J.A,Bello A.O,Volume 6, Issue 4, April 2016..
[2] The A Short Review, Bio-house Journal of Computer Science, Bio-house volume M Shiva Prasad, Asst. Professor. Issue 1, Jan-Feb 2016
[3] A Survey of Data Mining Techniques for Analyzing Crime Patterns, Ubon Thongsata pornwatana Defence Technology Institute Nonthaburi, 2016 IEEE
[4] Analysis and Prediction of Crimes by Clustering and Classification RasoulKiani ,Siamak Mahdavi, International Journal of Advanced Research in Artificial Intelligence,vol 4, No.8, 2015
[5] Crime Analysis and Prediction Using Data MiningShiju Sathyadevan, Devan M.S, 2014 IEEE
[6] Cluster Analysis for Anomaly Detection in Accounting Data An Audit Approach, The
[7] International Journal of Digital Accounting ResearceSutapatThiprungsri Miklos  Vasarhelyi, Vol. 11, 2011
[8] A Proposed Framework for Analyzing Crime Data Set Using Decision Tree and SimpleK-Means Mining Algorithms KadhimB. Swadi al-Janabi, Journal of Kufa for Mathematics and Computer, Vol.1, No.3, may 2011
[9] Crime Pattern Detection Using Data Mining Shyam Varan Nath Florida Atlanticuniversity/ Oracle Corporation. Proceedings of the 2006 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology
[10] Cluster Analysis for Reducing City Crime Rates Adel Ali Alkhaibari, Long Island University, Brooklyn, NY, Student Member, IEEE.
[11] Mining Crime Data by Using New Similarity Measure Guangzhu Yu1, Shihuang Shao1, Bing Luo