

Cluster based Zoning of Crime Info

Lalitha Saroja Thota¹,

Mohrah Alalyan¹, AL-Otaibi Awatif Khalid¹,

Fabiha Fathima¹,

¹College of Arts and Science, King Khalid University,
Khamis Mushayt, Abha, Saudi Arabia

Suresh Babu Changalasetty²,

Mohammad Shiblee²

²College of Computer Science, King Khalid University,
Abha, Saudi Arabia

Abstract— The criminal behavior is a disorderliness that is a combined result of social and economic aspects. The crime rate has expanded and the activities of criminals have broaden in last few decades due to better communication system and transport. Crimes cause terror and damage our community enormously in several means. In cities and towns the crime trends rises due to fast developmental activities and increase in population. In India, the regional location has a powerful impact on criminal activity. The CrimeInfo report of National Crime Records Bureau (NCRB), India collects, analyze and publish the crime data. The crime profiling and zoning can be modeled with utilization of data mining. In this paper, we make cluster analysis by using k-means cluster algorithm on criminal dataset of India. The cluster input is used to create custom India map with the cluster zones of states. The custom maps displays an overall crime profiles of states which helps police and law enforcement department to take additional preventive measures to combat against the 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 actions.

Keywords— Data mining, crime profiling, clustering, k-means

I. INTRODUCTION

In current era, criminals have maximal utilization of all modernized innovations and novel practices in perpetrating the crimes. Worldwide top priority is given by all government departments towards security and curtails the crime occurrence. A crime trend is a continuous, long-term rise or fall in temporally-based information. The crime records assumes a critical part in the planning of police working for control and discovery of crime. The Indian Police, throughout the years, have tried to enhance the proficiency of the crime records systems to perform their duties with higher productivity and adequacy. The automation of criminal records and the Police Computer Network system has delivers huge criminal information. The National Crime Records Bureau NCRB [1], home affairs ministry, India collects & maintains criminal data and publish reports of crime statistics documents. The crime data could be analyzed to interpret the emerging crime trends at high quality both locally and nationally.

The police department, crime researchers, criminologists and judicial officials in India use NCRBs vast statistical data to analyze and help curbing the crime. To execute crime analysis from huge criminal data, an appropriate scientific field

needs to choose. Data mining introduce drilling or deriving knowledge from historical huge database. Crime analysis is an attractive area where data mining shows a vital role in terms of investigation and forecasting. But the challenges to analyze the crime profiles and policing strategies is becoming more severe as the crime rate is rising day by day. In this research, we use data mining strategies on huge criminal dataset and knowledge gained is valuable and helps police department.

In this paper we use cluster technique of k-means algorithm on criminal dataset of India for crime analysis. The crime dataset is developed by applying complex query on the CrimeInfo India database [1]. The crime dataset is inputted into WEKA software to construct cluster zones based on k means clustering method. The cluster technique builds a model of states with high, medium and low crime zones. The cluster output of WEKA is passed manually as input to MyCustom map [2], an online interactive map tool of maps of India to create custom India map with the cluster zones of states. The custom maps displays an overall crime profiles of states which helps police and law enforcement department to take additional preventive measures 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.

II. BACKGROUND

The duty of criminal investigators at police departments all over the world is to identify and make tactical analysis of crime trends and patterns [3]. This analysis helps police to control the crime and enhance the safety of communities. In India NCRB [1] maintains the crime data and statistical publications software for analysis of crime info.

Shiva Prasad [4] discussed a short review of several Data Mining Techniques. An efficient comparing of various techniques of classification, especially Genetic Algorithm, K-Nearest Neighbor (kNN) and Support Vector Machines (SVM) of data mining was made to correlate the effects of the classification rules from data [5]. Data mining can be used to model crime profiling. Zakir Hussain, et al. [6] presented criminal behavior investigation by using data mining methods. The study elaborated the criminal behavior analysis of the offenders by using data mining approaches.

Manish Gupta et. al. [7] study existing system in use by Indian police and introduce a criminal analysis tool which is based on interactive queries. The tool helps police department to control crimes.

Priyanka Gera and Rajan Vohra [8] looks at the use of cluster technique in data mining to analyze the crimes patterns. The authors use k-means clustering to assist in the process of criminal profiling to primary crime data from Delhi police first information report (FIR) records. The study is used to estimate which type of crime is predominant and find which area categories are highly sensitive and show the distributions of each crime type in every area category

A comprehensive study on clustering strategies and its part on criminal applications was presented by Revatthy Krishnamurthy and Satheesh Kumar [9]. This review additionally helps crime branch for enhanced classification of crimes and forecasting .

Adeyiga and Bello [10] inspect many clustering methods used for crime profiling and propose a fuzzy clustering technique to crime analysts to forecast and counter crimes.

Sutapat Thiprungsri and Miklos Vasarhelyi [11] explore the likelihood of using cluster method of auditing for automating fraud filtering, a new abnormality discovery methods in the wire transfer framework.

Kadhim Swadi Al-Janabi [12] proposed scheme for crime data analysis to classify and cluster crimes based on data mining algorithms of decision tree and K Means technique.

Nath et.al [13] utilized k-means cluster algorithm to identify criminal pattern to pace up the mechanism of crime solutions.

Rapid miner tool was used by Jyoti Agarwal et al [14] for making criminal analysis by k-means clustering approach on crime dataset based on spatial distribution of existing information and expectation of crime rate. The results of analysis are extraction of patterns of crime, identification and forecasting of crime.

A new software tool using several data mining algorithms was developed by Malathi and Santhosh Baboo [15] to help effective crime investigation and analysis by police department for Indian scenario. The authors also utilize semi-supervised learning approaches on huge crime records to discover rich knowledge to enhance the precision of prediction.

III. METHODOLOGY

The CrimeInfo ver1.0 is a database software [1] developed by NCRB, India with the cooperation of the UN system. The crime dataset is developed by applying complex query on the CrimeInfo, India database.

The dataset obtained from NCRB does not contain any class label. To make groups of sets of the raw data it is better to

use cluster method. Clusterization is a concept of data mining which groups objects into a category based on some attributes. WEKA software is a tool to analysis the historical data with a collection of machine learning and data mining algorithms [16].

The crime dataset is inputted into WEKA software to construct cluster zones using K-means cluster method. The K-means clustering [16] algorithm groups objects depending on characteristics. The groups are developed by minimizing sum of squares of distances between data and concerned group centroid. Fig 1 shows k-means clustering algorithm.

The cluster output of WEKA is passed manually as input to MyCustom map [2], an online interactive map tool of maps of India to create custom India map with the cluster zones of states. The methodology adopted for Crime Zoning of Crime Info is shown in Fig 2

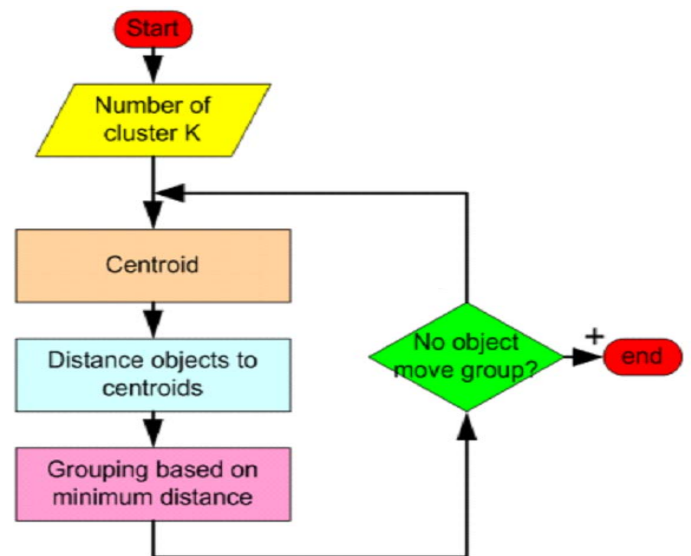


Fig. 1. K-means clustering algorithm

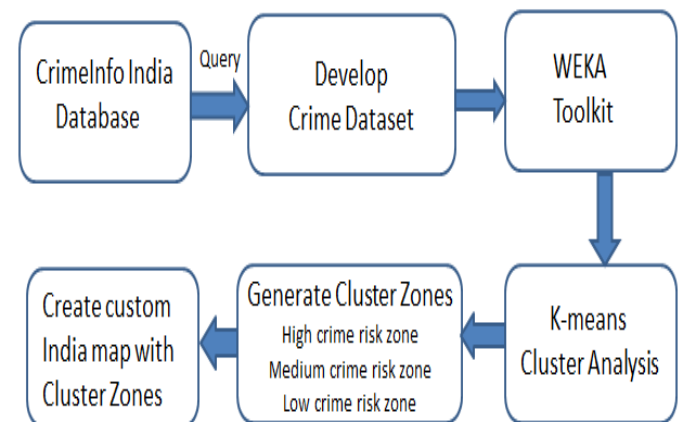


Fig. 2. Methodology for Crime Zoning

IV. EXPERIMENTS AND RESULTS

The crime dataset developed by applying complex query on the CrimeInfo India database [1] is shown in Table 1. The dataset contains state wise crimes made by male, female and total for the year 2010 in India.

TABLE 1: CRIME DATASET OF INDIA

State	Crimes (2010)		
	Female	Male	Total
Andhra Pradesh	16792	211813	228605
Arunachal Pradesh	32	2793	2825
Assam	751	69139	69890
Bihar	9338	186951	196289
Chhattisgarh	3438	69312	72750
Goa	238	3094	3332
Gujarat	13857	148186	162043
Haryana	2741	61496	64237
Himachal Pradesh	2451	16317	18768
Jammu & Kashmir	2319	29322	31641
Jharkhand	2161	49522	51683
Karnataka	12790	148828	161618
Kerala	6355	191157	197512
Madhya Pradesh	15587	327605	343192
Maharashtra	30118	275511	305629
Manipur	140	1166	1306
Meghalaya	28	1715	1743
Mizoram	78	2150	2228
Nagaland	65	1001	1066
Odisha	6387	83388	89775
Punjab	3719	45331	49050
Rajasthan	12032	165505	177537
Sikkim	18	928	946
Tamil Nadu	17112	194519	211631
Tripura	663	6172	6835
Uttar Pradesh	9127	282923	292050
Uttarakhand	533	12259	12792
West Bengal	11302	135293	146595
Andaman & Nicobar Islands	98	930	1028
Chandigarh	98	2485	2583
Dadra & Nagar Haveli	12	573	585
Daman & Diu	22	244	266
Delhi	1005	32493	33498
Lakshadweep	0	15	15
Puducherry	292	5287	5579

The dataset in table 1 is obtained in MS excel format. The raw dataset for WEKA is made from excel format by saving to CSV (comma delimited) format. The CSV data file is opened in WEKA tool to make clusters. Clusterization of data is done by selecting cluster option in main menu, choose simple K-means cluster type algorithm, give the attributes for clustering like distance function, number of clusters etc. Use the training data set as cluster mode and ignore the state attribute. Run the clustering by pressing start button. The WEKA output file with generated clusters and centroids with CrimeInfo dataset is shown in Fig 3.

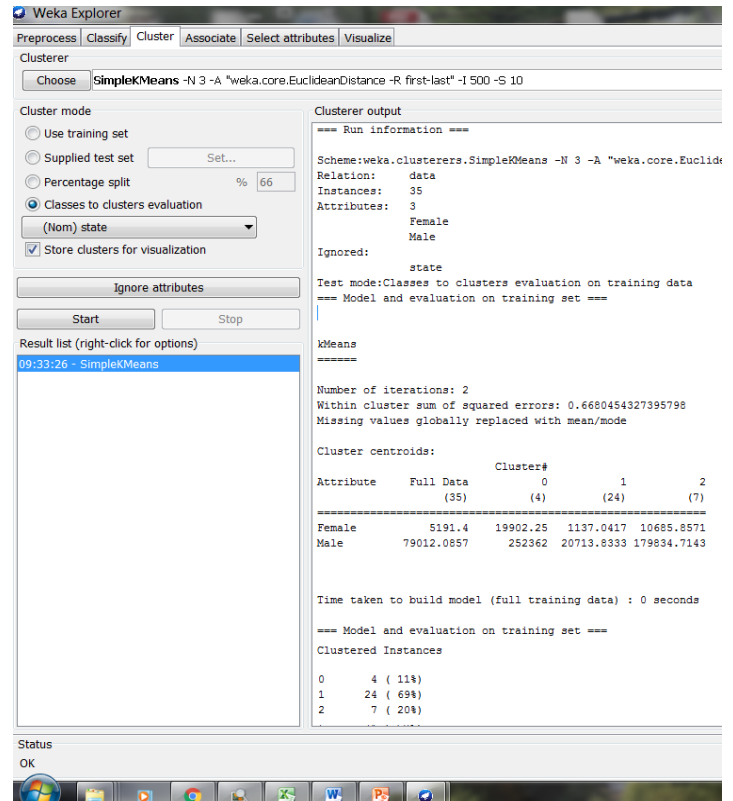


Fig. 3. WEKA output file with generated clusters and centroids with CrimeInfo dataset

Again run the clustering by using classes to cluster evaluation as cluster mode and ignore the state attribute. The WEKA output file with states clustered into crime zones is displayed in Fig 4. The cluster technique construct a model of states with high crime risk zone (cluster0), medium crime risk zone (cluster1) and low crime risk zone (cluster2) based on the total, male & female CrimeInfo are shown in Fig 5 to 7. The cluster output file of WEKA with states clustered into crime zones is passed as manual input to MyCustom map, an online interactive map tool of maps of India to create custom India map with the cluster zones of states is shown in Fig 8.

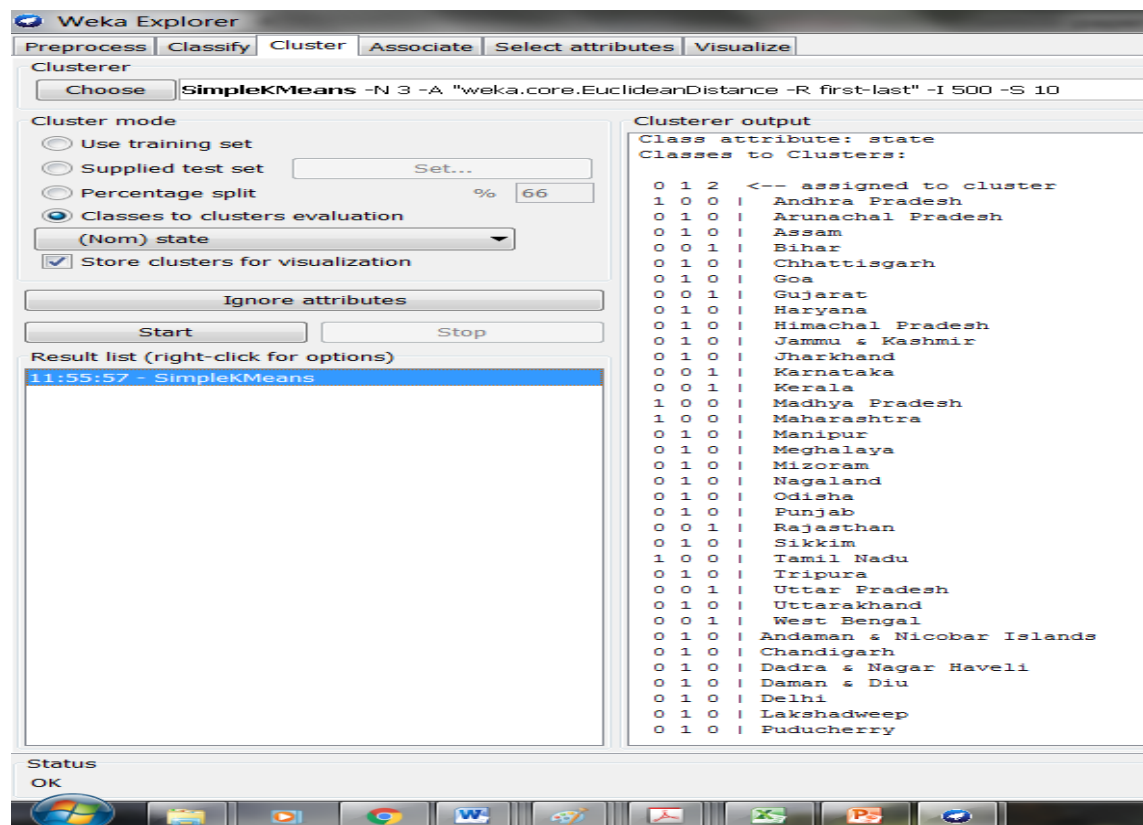


Fig. 4. WEKA Output file with states clustered into crime zones

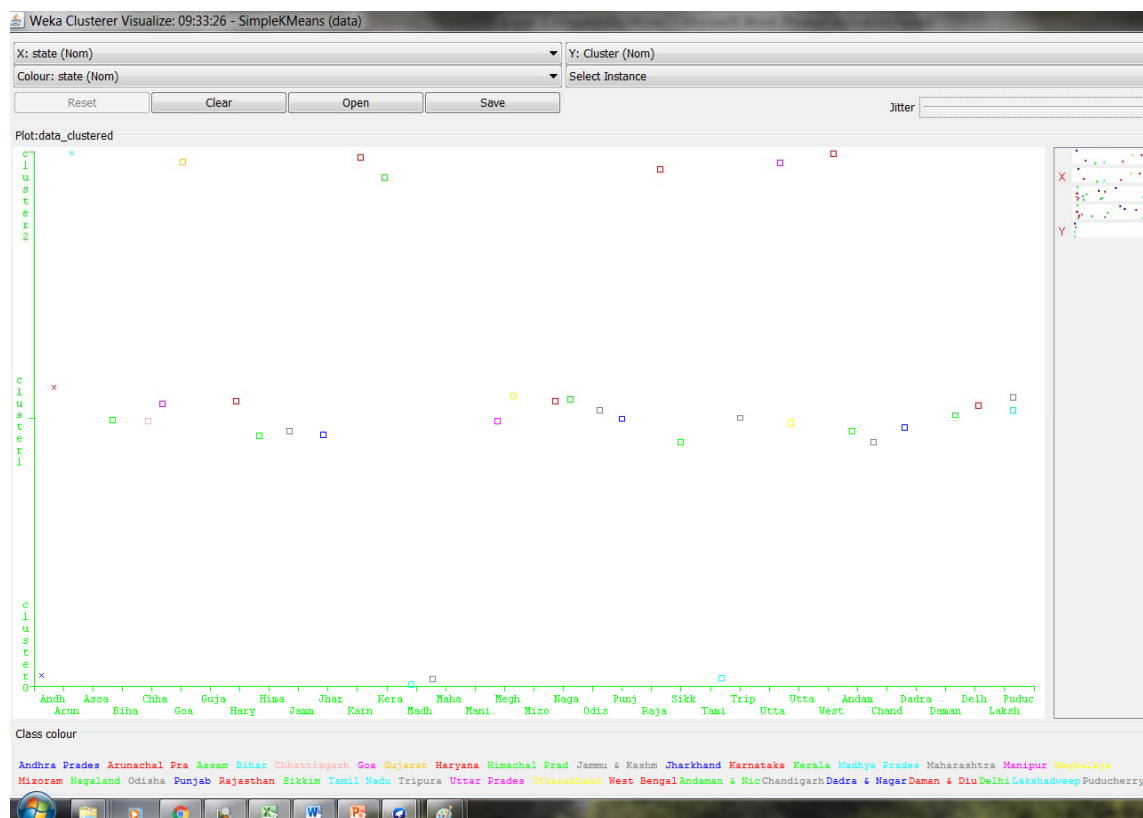


Fig. 5. States clustered into crime zones with total CrimeInfo

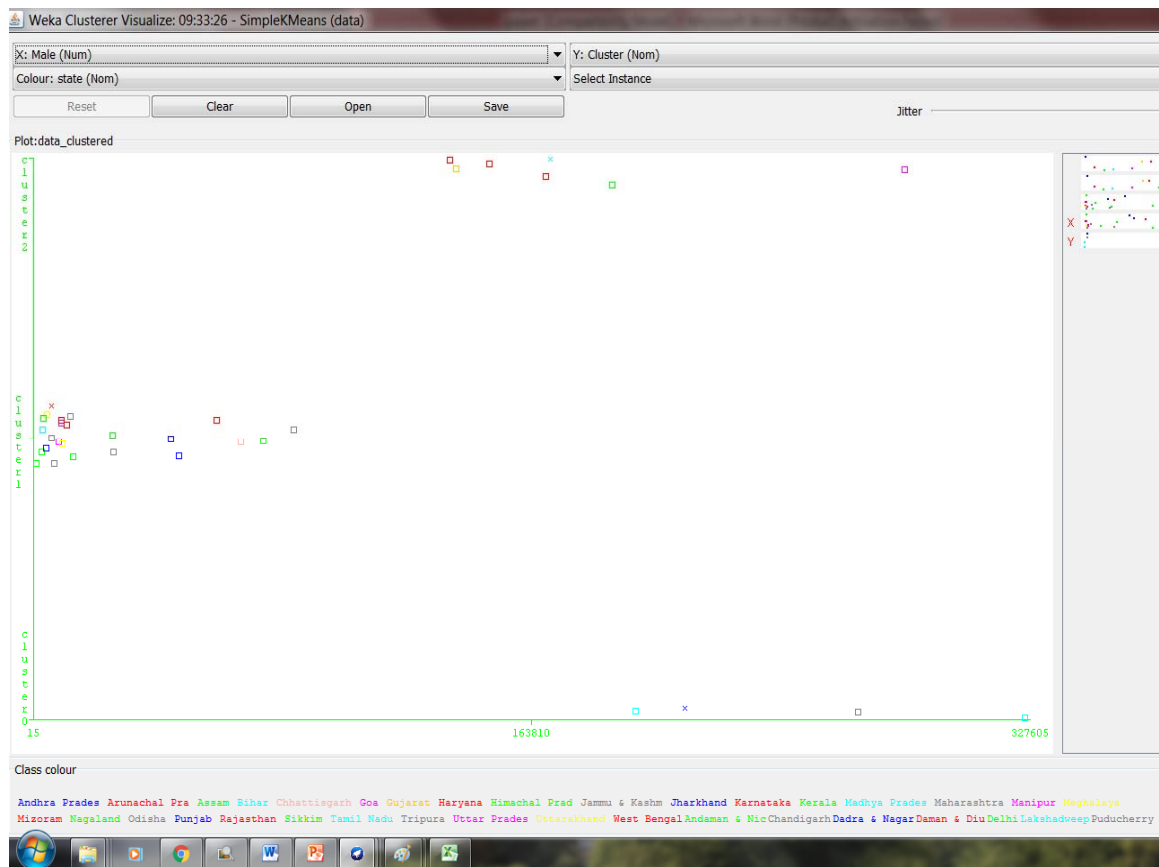


Fig. 6. States clustered into crime zones with male CrimeInfo

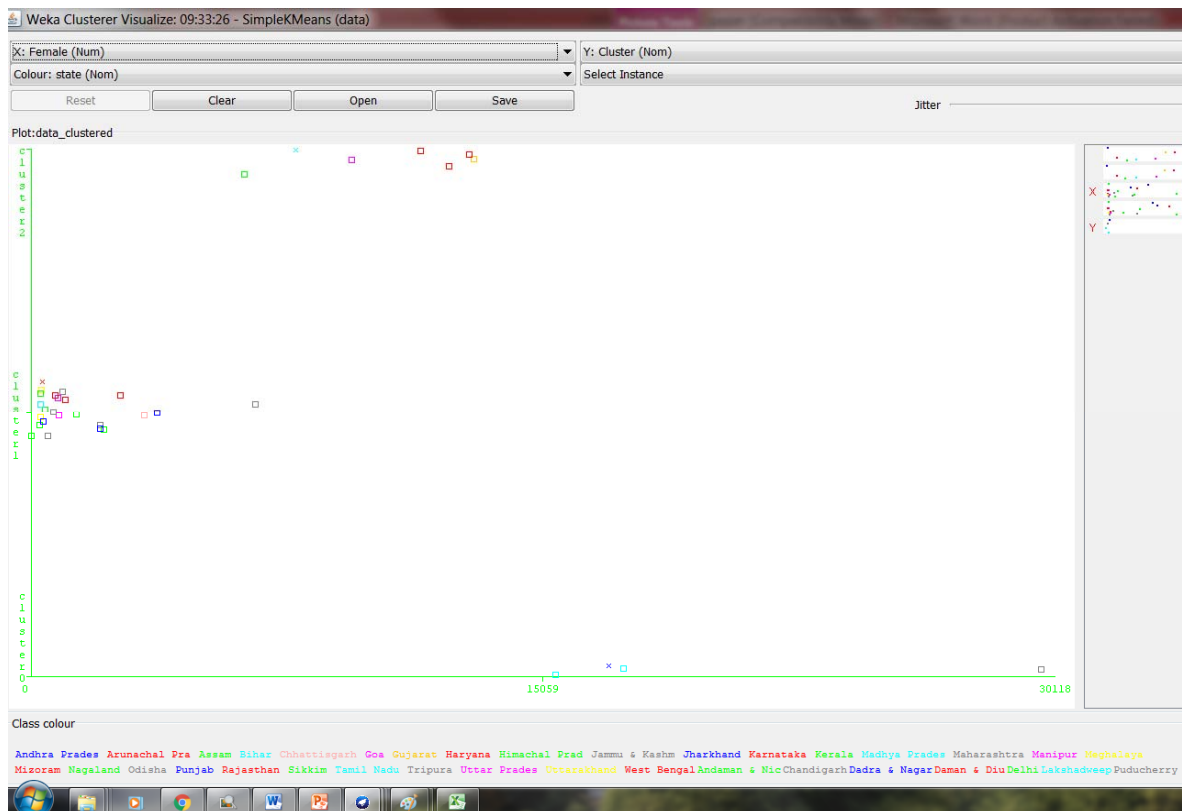


Fig. 7. States clustered into crime zones with female CrimeInfo

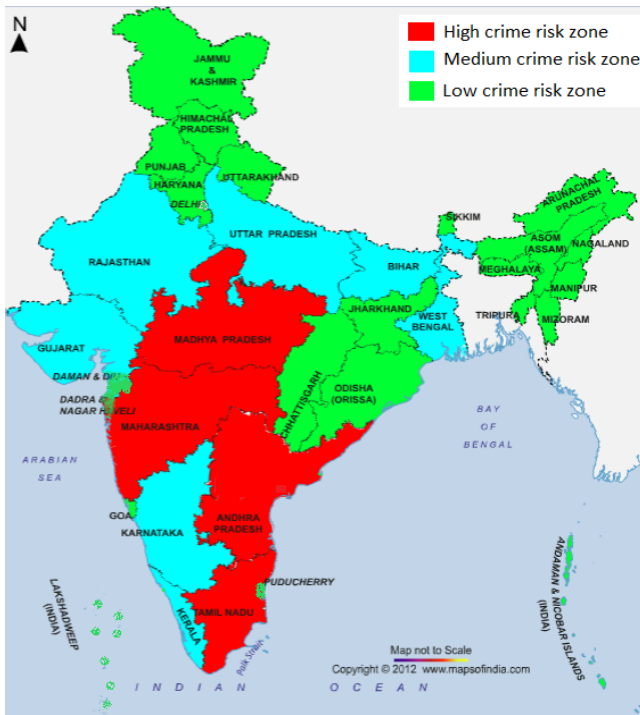


Fig. 8. Custom map of India with states clustered into crime zones

V. CONCLUSIONS AND FUTURE WORK

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 CrimeInfo NCRB dataset of India. The clustering method is used to group the states in India according to the criminal data of total, male and female number of crimes for the year 2010.

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.

REFERENCES

- [1] National crime records bureau, Ministry of home affairs, India, Software for data analysis, Crime Info (Crime in India) <http://ncrb.nic.in/index.htm>
- [2] MyCustom map tool, <http://www.mapsofindia.com/editable-maps/customise-your-map.html>
- [3] Chris Delaney, Crime Pattern Definitions for Tactical Analysis, International Association of Crime Analysts (IACA), August 2011
- [4] M Shiva Prasad, Data Mining Techniques: A Short Review, Biohouse Journal of Computer Science, Biohouse volume 2 issue 1, Jan-Feb 2016
- [5] Dr. J.Subash Chandra Bose, Dr. Suresh Babu Changalasetty et al, A Hybrid GA/kNN/SVM Algorithm for Classification of data, Biohouse Journal of Computer Science, Biohouse volume 2 issue 2, Mar-Apr 2016
- [6] K. Zakhir Hussain, M. Durairaj and G. Rabia lahani Farzana ,*Criminal behavior analysis by using data mining techniques*, IEEE International Conference On Advances In Engineering, Science And Management (ICAESM) March , 2012
- [7] Manish Gupta, B.Chandra and M. P. Gupta, Crime Data Mining for Indian Police Information System, Computer society of India, 2007
- [8] Priyanka Gera and Rajan Vohra, City Crime Profiling Using Cluster Analysis, International Journal of Computer Science and Information Technologies IJCSIT, Vol. 5 (4), 2014
- [9] Revatthy Krishnamurthy and J. Satheesh Kumar, Survey of data mining techniques on crime data analysis, International Journal of Data Mining Techniques and Applications, Vol 01, Issue 02, December 2012
- [10] Adeyiga J.A and Bello A.O, A Review of Different Clustering Techniques in Criminal Profiling, International Journal of Advanced Research in Computer Science and Software Engineering IJARCSSE, Volume 6, Issue 4, April 2016
- [11] Sutapat Thiprungsri, Miklos A. Vasarhelyi, Cluster Analysis for Anomaly Detection in Accounting Data : An Audit Approach, The International Journal of Digital Accounting Research Vol. 11, 2011
- [12] Kadhim B.Swadi al-Janabi . A Proposed Framework for Analyzing Crime DataSet using Decision Tree and Simple K-means Mining Algorithms, Journal of Kufa for Mathematics and Computer, Vol.1, No.3, may 2011
- [13] S.V. Nath, Crime pattern detection using data mining, IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology, 2006
- [14] Jyoti Agarwal, Renuka Nagpal, Rajni Sehgal, Crime Analysis using K-Means Clustering, International Journal of Computer Applications IJCA, Volume 83 – No4, December 2013
- [15] A. Malathi, Dr. S. Santhosh Baboo, Algorithmic Crime Prediction Model Based on the Analysis of Crime Clusters, Global Journal of Computer Science and Technology Volume 11 Issue 11 Version 1.0 July 2011
- [16] Suresh Babu Changalasetty, Lalitha Saroja Thota, Ahmed Said Badawy and Wade Ghribi, Classification of Moving Vehicles using K-Means Clustering, IEEE International Conference on Electrical, Computer and Communication Technologies (ICECCT-2015), India, Mar 2015