Expert Recommendation for Crime Prediction Using Data Mining Techniques

[1]M. Sethupathi, [2]G. Sai srikar, [3]R. R. Sundar Naresh, [4]Dr. R. Priyatharshini

[1,2,3]Student, [4]Professor
Department Of Information Technology,
Easwari Engineering College, Chennai, India

_____

Abstract: **Data mining is the process of extracting essential and required information and discovering patterns from a large amount of data. Crime analysis is a systematic way to deal with distinguishing and examining trends and patterns in crime. With the expanding of automated frameworks, crime information investigators can help the Law requirement officers to accelerate the way toward illuminating violations. The crime data analysts can help the Law implementation officers to accelerate the way toward settling crimes. The proposed framework can anticipate locales which have a high likelihood of crime event and can imagine crime inclined zones. Rather than concentrating on cause or reason of crime event, we are concentrating principally on crime elements of every day. Based on the area recommended a Better officer can be recommended.**

IndexTerms - **K-Means clustering, Robust Clustering, Case-based Reasoning, Crime analysis, Recommendation system, Predictive analysis, Crime-patterns, law-enforcement.**
_____

I. INTRODUCTION

A crime rate has become a major topic of concern certainly to limit the development of good governance and increasing day by day. Crime prediction and criminal identification are the major problems to the police department as there are voluminous data of crime exist. So, we need methodologies to predict and prevent crime. Data Mining provides clustering and classification technique for this purpose. Clustering is used for grouping the similar patterns to identify crimes. Cluster refers to a geographical collection of crime that can be visualized using the geo-spatial plot in the map.

When crimes like robbery, firebombing etc. have been decreased, crimes like murder, kidnap etc. have been increased. The crime rates accelerate continuously and the crime patterns are constantly changing. As a result, the behaviors in crime pattern are difficult to explain. We cannot analyze the victims of crime but can analyze the place where the crime occurred or happened.

The perfect results cannot be generated yet the outcomes demonstrate that our application helps in diminishing the crime rate to a specific degree for building such a capable crime investigation apparatus we need to gather crime records and assess it. Since the availability of criminal data or records is limited we are collecting crime data from various sources like websites, news sites, blogs, social media, RSS feeds etc. This huge data is used as a record for creating a criminal record database. So, the main challenge in front of us is developing a better, efficient crime pattern detection tool to identify crime patterns effectively. The main challenges we are facing are:
  • Increase in crime data that must be put away and investigated.
  • Analysis of information is troublesome since information is inadequate and conflicting.
  • Limitation in getting crime information records from Law Enforcement division.

The reason for choosing this method is that we have only data about the known crimes we will get the crime pattern for a particular place. Thus, classification technique that will depend on the current and known settled violations, won't give great prescient quality for future wrongdoings. Additionally, nature of wrongdoings changes after some time, so with a specific end goal to have the capacity to recognize more up to date and obscure patterns in future, clustering techniques work better. The objective of this work is to predict crime by using clustering methods.

II. LITERATURE SURVEY

Malathi and Dr. S. Santhosh Baboo (2011) built up a crime analysis tool for Indian situation by utilizing different information mining strategies like k-means clustering and DBSCAN clustering. The authors conclude that the tool can be used by the Indian police and law agencies for crime detection and prevention as it provides faster analysis results and identifies common crime patterns. In future, the system should be trained in various scenarios.

Mugdha Sharma et al. proposed a progressed ID3 calculation for exhibiting significance characteristic hugeness on the traits which have less esteems yet higher significance, instead of the qualities with more esteems and lower significance and also tackle the characterization imperfection to pick attributions with more esteems. The investigation of the exploratory information demonstrates that the progressed ID3 calculation gets more sensible and more viable characterization rules. The system lags in different methods for feature selection and classification.

Navjot Kaur et al., (2012) presented an overview of k-means clustering. They used ranking method for k-means clustering and compared its performance with the traditional k-means clustering. As a result of the comparison, the authors conclude that the ranking based k-means clustering had less execution time and provide better results than the traditional method.

SushantBharti et al. proposed hidden link algorithm to detect hidden links of the networks of co-offenders which show the possible future crime partner and different network beyond the real network. This paper also analyzes the centrality of a node. This analysis describes the importance of node of the network. This is used to discover the strongest person, the power of the person and role of the person in the network. This paper gave future approaches i.e. predictive approach in crime analysis which helps in stopping the crime before it occurs and also analyze the network of Co-offenders in India and predict the possible future network of offenders. The future work will be used the predictive approach in crime analysis which helps in stopping the crime before it occurs.

ShijuSathyadevan et al. proposed Apriori algorithm to identify the trends and patterns in crime. This algorithm is also used to determine association rules highlighting general trends in the database. This paper has also proposed the naïve Bayes algorithm to create the model by training crime data. After testing, the result showed that Naive Bayes algorithm gave 90%accuracy. In future, It will be more accurate if we consider a particular state/region. Also, another problem is that we are not predicting the time in which the crime is happening. Since time is an important factor in crime we have to predict not only the crime-prone regions but also the proper time.

The conclusion of the survey is even though these systems can handle crime data but accuracy is very less compared to our system whereas our system can tell the probability about the next crime that will take place.

III. PROPOSED SYSTEM

The proposed system will analyze previous year crime data records and will be able to detect frequent crime areas. By using unsupervised learning methodology such as K-Means Clustering Algorithm & Robust clustering algorithm we will able to predict the major crime occurrences in a city. We are able to precisely cluster the geospatial points so that the hotspot is of reasonable size. Utilizing kernel density estimation(KDE) Algorithm we can appraise the likelihood density capacity of an arbitrary variable, where it will demonstrate what is the likelihood that the will occur in a specific region with the assistance of Case-Based Reasoning, the framework will suggest an Expert in view of the sort and recurrence of violations in the specific city.
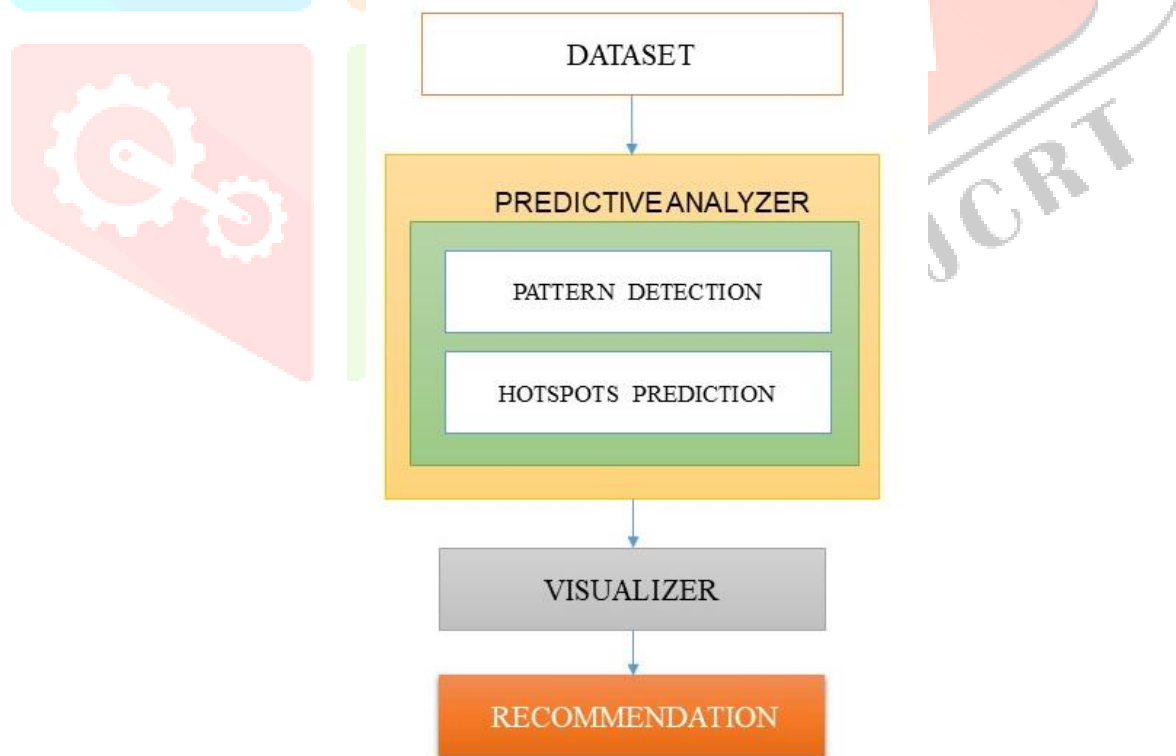

Fig. 1. The architecture of the Proposed system

The crime attributes are extracted from crime dataset such as types of crime, Latitudes, Longitudes etc., The Crime patterns are detected using k-means clustering and robust clustering. The probability of the crimes is calculated using KDE (Kernel Density Estimation) the Geospatial points being visualized in the crime spots in the certain area. The special officer is recommended based on crime data using case-based reasoning and police officer database.

IV. METHODOLOGY

A. Data Acquisition Module

The data is acquired from various databases from government website stored in a table format which are used to predict crime pattern. The dataset covers the years 2008-2014 and contains records translated from the first materials distributed by the Government of India, Ministry of Home Affairs. The dataset (Past Crime Records) are imported into the application in a CSV file format. The latitude and longitude data are extracted from the given data set. The dataset contains many variables such as data Id, Cluster Number, Date, Time, Time Category, Day, Victim Age, Age Category, Gender, Vehicle, Weapon, Place, Place Category, Address, Police Station, Location Type, Theft Value Rupees, Theft Category, Number of Suspects, Latitude, Longitude and Involvement of Same Gang Suspected. Using these attributes, the crime analysis is made and predicted

	data Id	Cluster	Date	Time	Time Cate	Day	Victim Age	Age Categ	Gender	Vehicle	Weap	Place	Place Cate	Address	Police Sta	Location T	Theft Va	Theft Cat	Num	Latitude	Longitude	Inv
1																						
2	0	1	22/10/2008	2:30 PM	Afternoon	Wednesda	73	Very Old	F		0	0	Shukrawar Peth	Lower Mic	Swami Sar	Vishramb	Street	8000	Low	1	18.51104	73.86381	A w
3	1	2	23/10/2008	2:30 PM	Afternoon	Thursday	0	0	M		0	0	Swargate	Lower Mic	Satara roa	Swargate	Street	36500	Low	4	18.50002	73.85857	two
4	2	outlie	23/10/2008	8.30 PM	Night	Thursday	21	Young	F	Motorcycle	0	Sahakarnagar	Lower Mic	Santnagar	Sahakarna	Street	10000	Low	2	18.48438	73.85349	
5	3	2	10/1/2009	8:00 PM	Night	Saturday	52	Old	F	Motorcycle	0	Bhusari Colony	Lower Mic	Bhusari Co	Kothrud	Street	33000	Low	3	18.50839	73.79035	pre
6	4	3	11/2/2009	2:30 PM	Afternoon	Wednesda	45	Middle Ag	F	Motorcycle	0	Pimpri	Lower mic	Sant Tukai	Pimpri	Street	30000	Low	2	18.62316	73.81985	
7	5	2	11/2/2009	3:30 PM	Afternoon	Wednesda	62	Very Old	F		0	0	Dhankawadi	Poor	Atharva re	Sahakarna	Street	20000	Low	3	18.52398	73.85346	imp
8	6	4	14/03/2009	8:30 AM	Morning	Saturday	77	Very Old	F	Motorcycle	0	Dattawadi	Poor	Dattawadi	Dattawadi	Street	30000	Low	2	18.65328	73.77794	
9	7	4	14/03/2009	9:30 AM	Morning	Saturday	58	Old	F	Motorcycle	0	Sadashiv Peth	Lower mic	Subhashn	Sadashiv F	Street	30000	Low	2	18.50474	73.85735	go
10	8	4	14/03/2009	7:00 PM	Evening	Saturday	39	Adult	F	Motorcycle	0	Yerawada	Lower Mic	Yerawada	Vishrantw	Street	19000	Low	2	18.55841	73.88601	
11	9	5	16/04/2009	2:30 PM	Afternoon	Thursday	40	Adult	F	Motorcycle	0	Thergaon	Lower mic	near Birla	Hinjewad	Street	29000	Low	1	18.61589	73.76554	Mu
12	10	5	17/06/2009	8:45 AM	Morning	Wednesda	44	Middle Ag	F		0	0	Pune-Alandi Roa	Lower mic	Thorlya Pa	Vishrantw	temple	12000	Low	1	18.54548	73.92214	tak
13	11	6	17/06/2009	9:30 AM	Morning	Wednesda	0	0	F		0	0	Dighi	Poor	Dighi	Dighi	Street	45000	Medium	1	18.61242	73.87721	
14	12	6	17/06/2009	2:30 PM	Afternoon	Wednesda	0	0	F		0	0	Dighi	Poor	Dighi	Dighi	Street	50000	Medium	1	18.61242	73.87721	
15	13	2	25/07/2009	8:30 PM	Night	Saturday	60	Old	F	Motorcycle	0	Model Colony	Upper mic	Model Co	Model Co	Street	16000	Low	2	18.52828	73.84174	ret
16	14	7	26/07/2009	7:30 AM	Early Mori	Sunday	56	Old	F	Motorcycle	0	Marketyard	Lower mic	Gangadha	Marketyar	Temple	19000	Low	2	18.48358	73.87144	
17	15	7	26/07/2009	8:00 AM	Early Mori	Sunday	52	Old	F	Motorcycle	0	Pimple Saudagar	Lower mic	Pimple Sa	Sangvi	House	36000	Low	2	18.5967	73.79779	
18	16	7	1/8/2009	7:00 PM	Evening	Saturday	49	Middle Ag	F	Motorcycle	0	Katraj	Lower mic	Katraj	Sahakarna	Street	36000	Low	2	18.44159	73.86837	on
19	17	2	18/08/2009	8:45 AM	Morning	Tuesday	54	Old	F		0	0	Nigdi	Lower mic	Nigdi	Nigri	Street	15000	Low	1	18.65877	73.77759	pre
20	18	2	10/10/2009	3:30 PM	Afternoon	Sunday	23	Young	F	Motorcycle	0	Hadapsar	Lower Mic	Hadapsar	Wanowrie	Street	7500	Low	1	18.50661	73.91979	
21	19	8	10/10/2009	7:00 PM	Evening	Sunday	46	Middle Ag	F		0	0	Khilarwadi	Poor	Khilarwad	Deccan Gy	Street	70000	Medium	1	18.51676	73.84296	
22	20	1	10/10/2009	7:00 PM	Evening	Sunday	84	Very Old	F		0	0	Wakdewadi	Lower mic	Wakdewa	Khadki	Street	10000	Low	2	18.53859	73.85059	
23	21	2	18/10/2009	9:00 PM	Night	Sunday	42	Middle Ag	F	Motorcycle	0	Sahakarnagar	Lower mic	Sahakarna	Sahakarna	Street	30000	Low		18.48668	73.84448	

Table I. Crime Dataset

B. Hotspot Detection Module

The hotspot is detected using the process of K-means clustering of the geo-spatial labels and marking the only region of high crime rates. The hotspot is identified and marked by learning from past crime patterns. Using the robust clustering the number of clusters can be reduced.

The paper tends to help an expert in finding designs, patterns, making conjectures, discovering connections and conceivable clarifications, mapping criminal systems and recognizing conceivable suspects.
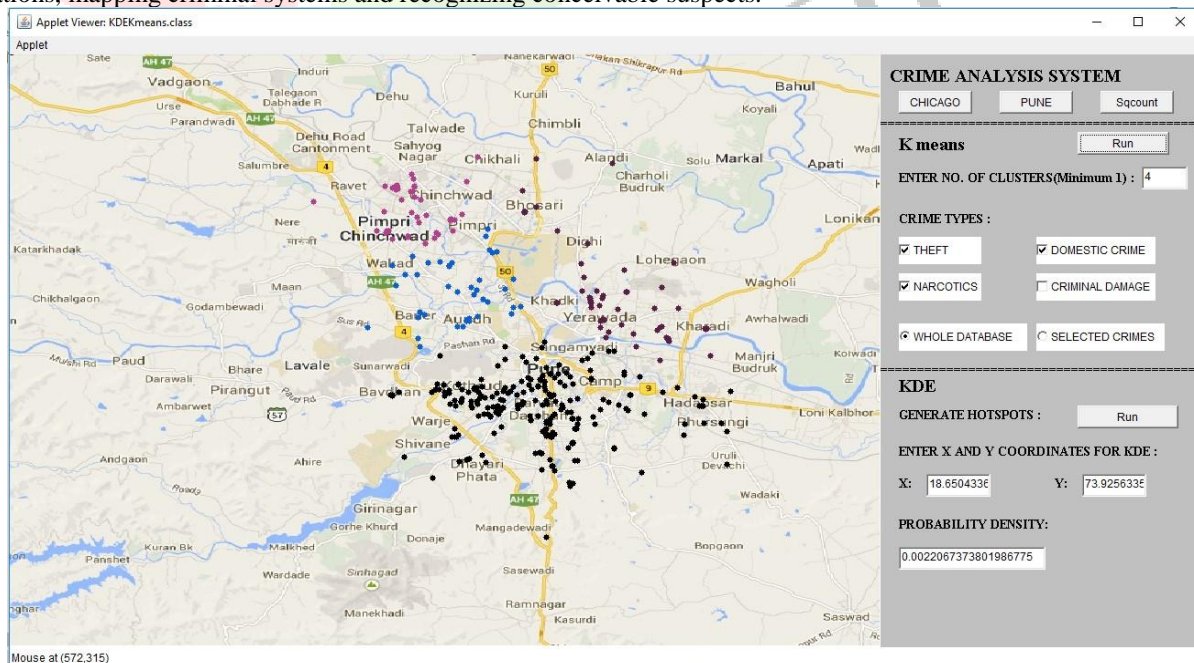

Fig. 2. Map of Pune City about Geo spatial Mapping

C. Kernel Density Estimation (KDE)

Using Kernel Density Estimation, the probability of the crime to happen next will be predicted based on the past criminal record. The crime is filtered based on the type of crime, then using the coordinates the points are mapped, the KDE algorithm is used to find kernel density estimation.
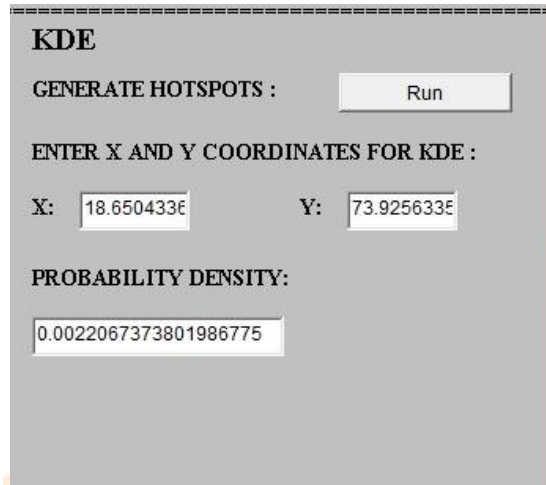

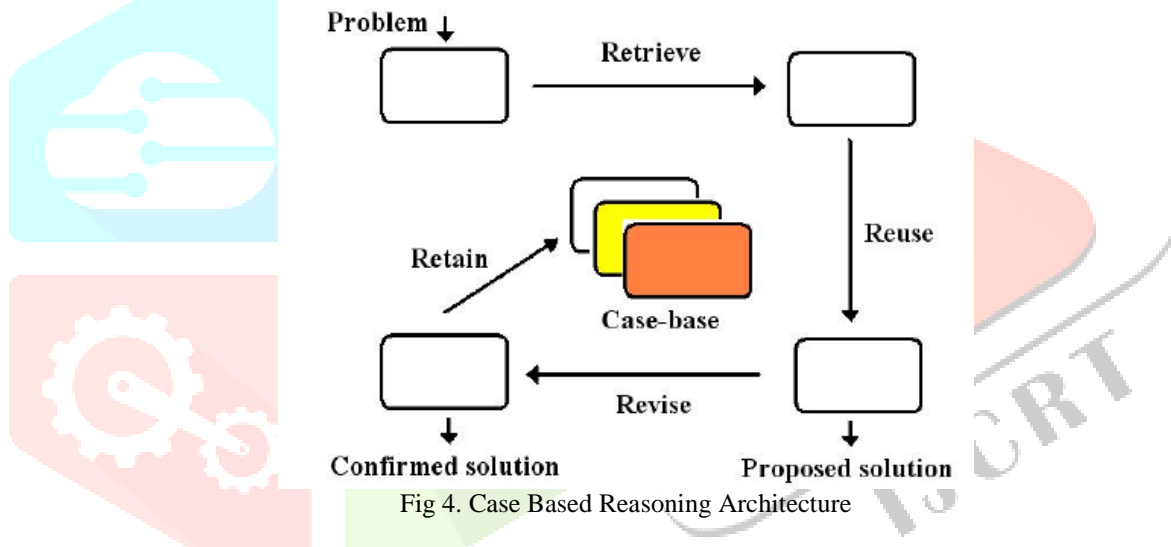Fig 3. Kernel Density Estimation Module

D. Expert Recommendation Module

Fig 4. Case Based Reasoning Architecture

The final results are displayed in the form of the major hotspot region which includes a trend of criminal activities. This can be used to recommend expert for a particular city or region based on the crime pattern. Case-based reasoning is used for recommending experts using police officer database. Every police officer has their own weightage based on the previous cases they handled. Based on this the police officers are recommended.
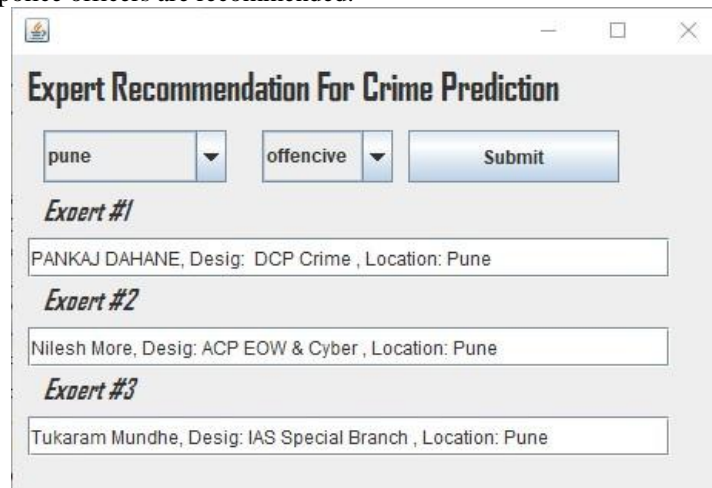

Fig 5. Expert Recommendation System

V. **Experimental Results**

The crime attributes such as latitude, longitude, type of crime and no of clusters are taken from the dataset. The attributes from the dataset are taken as input. Computation time is the length of time taken to compute the dataset and clustering of the data. The error rate can be considered as an accuracy rate where it differs from the expected output.

Parameters	K-Means	ID3
No of Clusters	6	6
Error Rate	0.3395	0.3580
Computation Time	16ms	156ms

Table II:Comparative Performance Analysis of Proposed System and existing system

Fig 6. Computation Time

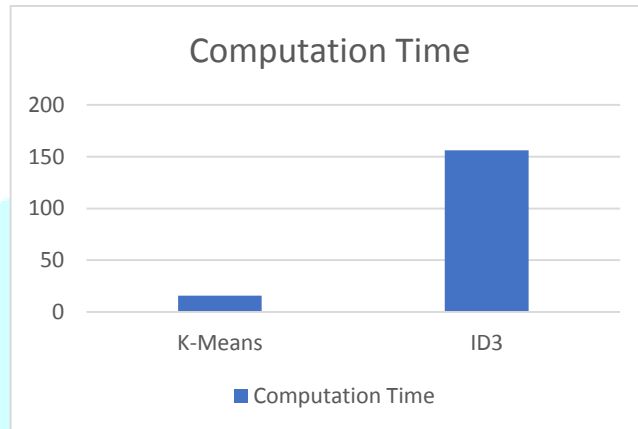

Fig 7. Computation Time
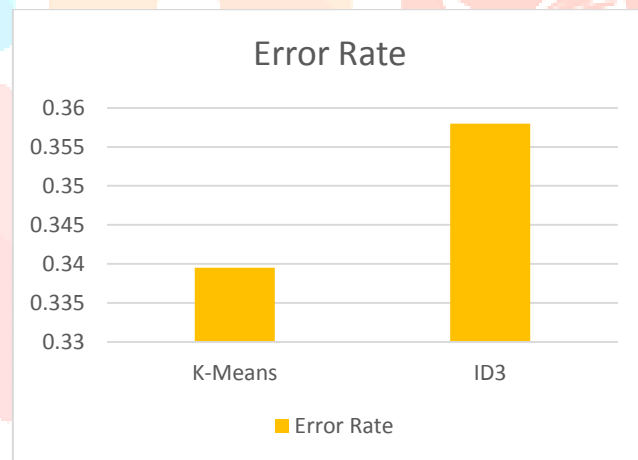
VI. CONCLUSION

In this paper will prediction and recommendation based on different test sets. Clustering in view of k-means clustering with over 80% exactness and utilized the Kernel Density Estimation (KDE). Our system takes factors/attributes of a place and type of crime gives the frequent patterns of that place. The pattern is used for building a prediction model. We used Cased Based Reasoning for Recommending an expert police officer.

Till now we trained our system using certain attributes but we are planning to include more factors to improve accuracy. Our software predicts crime-prone regions in India on a certain probability. It will be more accurate if we consider a particular state/region. Also, another problem is that we are not predicting the time in which the crime is happening. Since time is an important factor in crime we have to predict not only the crime-prone regions but also the proper time.

REFERENCES

[1] ShijuSathyadevan, Devan M.S and Surya Gangadharan.S. "Crime Analysis and Prediction Using Data Mining", First International Conference on Networks & Soft Computing, 2014.
[2] The use of predictive analysis in spatiotemporal crime forecasting: Building and testing a model in an urban context - https://doi.org/10.1016/j.apgeog.2017.06.011

[3] Mugdha Sharma. "Z - CRIME: A Data Mining Tool for the Detection of Suspicious Criminal Activities Based on Decision Tree", International Conference on Data Mining and Intelligent Computing, 5-6 September 2014.

[4] SushantBharti, Ashutosh Mishra. "Prediction of Future possible offender's network and role of offender's", Fifth International Conference on Advances in Computing and Communications, 2015.

[5] Prashant K. Khobragade and Latesh G. Malik."Data Generation and Analysis for Digital Forensic Application using Data mining", Fourth International Conference on Communication Systems and Network Technologies, 2014.

[6] K.ZakirHussain, M. Durairaj and G. RabialahaniFarzana. "Criminal Behavior Analysis By Using Data Mining Techniques", IEEE-International Conference on Advances in Engineering, Science and Management (ICAESM -2012), March 30-31, 2012.

[7] Mohammed MahmoodAli, KhajaMoizuddin Mohammed and Lakshmi Rajamani."Framework for Surveillance of Instant Messages in Instant messengers and Social neworking sites using Data Mining and Ontology", IEEE- Students' Technology Symposium, 2014.

[8] A. Q. Ansari, TapasyaPatki, A. B. Patki and V. Kumar. "Integrating Fuzzy Logic and Data Mining: Impact on Cyber Security ", IEEE-Fourth International Conference on Fuzzy Systems and Knowledge Discovery, 2007.

[9] B. Chandra, Manish Gupta and M. P. Gupta. "A Multivariate Time Series Clustering Approach for Crime Trends Prediction", IEEE International Conference on Systems, Man and Cybernetics, 2008.

[10] Rajesh Kumar, Nikhil R. Pal, BhabatoshChanda and J. D. Sharma."Detection of Fraudulent Alterations in BallPoint Pen Strokes using Support Vector Machines", IEEE India Conference(INDICON), 2009.

[11] Veena H Bhat, Prasanth G Rao, Abhilash R V, DeepaShenoy, Venugopal K R and L M Patnaik. "A Novel Data Generation Approach for Digital Forensic Application in Data Mining", IEEE Second International Conference on Machine Learning and Computing,2010.

[12] T.Subbulakshmi, Dr. S.Mercy Shalinie, V. Ganapathi Subramanian, K.BalaKrishnan, D.AnandKumar and K.Kannathal ."Detection of DDoS Attacks using Enhanced SupportVector Machines with Real Time Generated Dataset", Third International Conference on Advanced Computing (ICoAC), 2011.

[13] MayankPandey and Vadlamani Ravi. "Detecting phishing e-mails using Text and Data mining", IEEE International Conference on Computational Intelligence and Computing Research, 2012.

[14] Vinit Kumar Gunjan, Amit Kumar and ShardaAvdhanam. "A Survey of Cyber Crime in India",15th International Conference on Advanced Computing Technologies (ICACT), 2013.

[15] Malathi. A and Dr. S. SanthoshBaboo. "An enhanced algorithm to predict a future crime using data mining", International Journal of Computer Applications, 1-6, May 2011.

[16] Crime dataset - https://data.gov.in/.