WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Crime Rate Analysis and Prediction Using K-Means

Srushti Umasare[1], Srushti Phirke[2], Sneha Thakur[3], Vaishali Kulkarni[4]

Department of Electronics and Telecommunication
JSPM's Rajarshi Shahu College of Engineering, Pune, India

ABSTRACT

Today, criminal activities are increasing worldwide due to modern technologies, which have an adverse effect on society and common people. The primary objective of this paper is to analyze the crime pattern and predict the crime occurrence rate in the future to take appropriate measures against it. One of the most suitable methods to apply to a huge amount of crime datasets is the data mining technique to organize and analyze the data. The k-means clustering algorithm is applied to form clusters based on crime in various regions to find generic patterns. A further prediction is done using the Random Forest Regression Model to get the high crime-prone areas.

Keywords:Crime dataset, data mining, k-means clustering, prediction, and Random Forest Regression Model

1. INTRODUCTION

In the current scenario, criminals are becoming more technologically advanced which serves as a major challenge for law agencies and government officials to analyze huge amounts of criminal datasets. So, the basic requirement is to select appropriate crime analysis and data mining techniques. A data mining algorithm can be used to get clusters of cities or areas having similar crime occurrence rates and analyze the crime pattern. In this paper,the k-means clustering technique of data mining is used to extract useful information from the huge crime dataset.Clustering will be done based on the place where the crime occurred. This will help to predict crimesthat will occur in the future. Which would help the police to identify the crime "hot spots" and take the required actions

2. LITERATURE SURVEY

[1] S.G. Krishnendu, P. P. Lakshmi, and L Nitha proposed a system to examine the possibility of using clustering technology for crime analysis. They found out which state has more or fewer criminals based on each cluster's values. The result can be used to make various strategies for crime control in the future.

[2] Gouri Jha et al. proposed a system for finding a pattern that is necessary for prediction and finding criminal patterns and behaviors. In the future, they tend to focus on examining fuzzy clustering algorithms to apply to serial killing-relateddatasets and examine the analysis of the same.

[3] Khushabu A. Bokde et al. proposed a project that focuses on crime analysis by considering crime homicide and plotting it concerning year. They concluded that homicide is decreasing from 2009 to 2018.

3. PROPOSED SYSTEM ARCHITECTURE

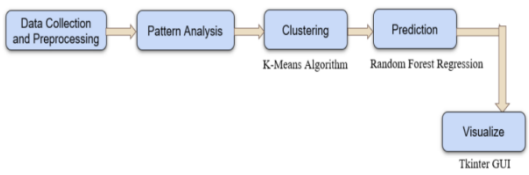


Fig 1: Block Diagram of proposed system Architecture

Figure 1 shows the block diagram of the proposed system architecture and below are the steps for implementation:
1.Crime dataset is taken.
2. One-hotlabel encoder is used to filter and pre-process the dataset.
3.Elbow method is used to get an effective number of clusters or centroids from the elbow graph.

4. K means clustering is performed on the resultant dataset and various clusters of cities were obtained.
5.Random Forest Regression Model is applied to the dataset and the type of crime in the future is predicted.

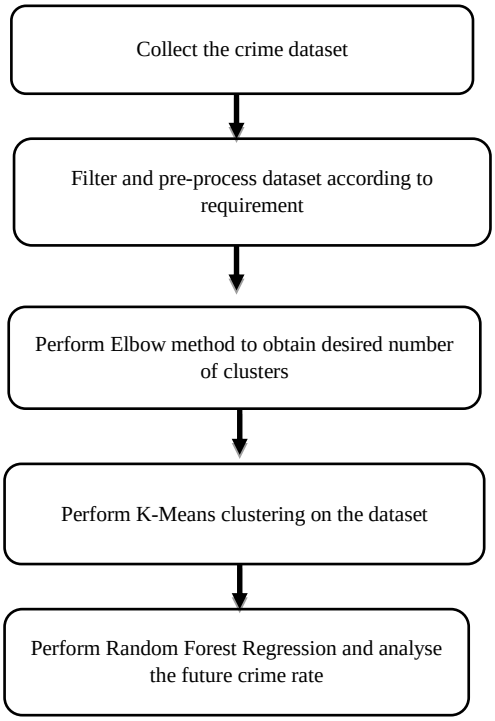


Fig 2: Implementation steps of system architecture

4. EXPERIMENTAL SETUP AND RESULT

4.1. Approach Used:

4.1.1 Data Pre-processing

The dataset is collected from *data.World* website. If the data is dirty, it will generate incorrect visualizations, hence leading to incorrect conclusions. The crime data collected also has some level of dirtiness such as null values, inconsistent date formats, and some outliers.
The Null or missing values are replaced by the mean values of the particular column or in the case of a huge dataset sometimes the entire row is deleted. To convert the categorical data into labels or numeric values we use the method called one-hot label encoding. In this way, the dataset is made meaningful for further process.

4.1.2 Data Analysis

From the Dataset present, we can analyze the crime pattern in a particular city or state with the help of various plots. It makes it easy for the officials to analyze the dataset to easily identify the most crime-prone area and take the necessary action.

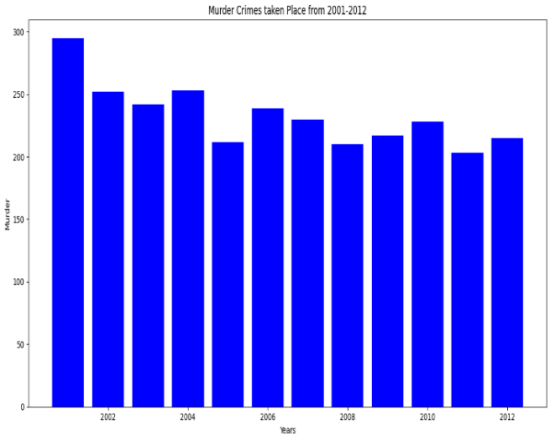


Fig 4: Plot of total no. of murders took place in the state from year 2001-to 2012.

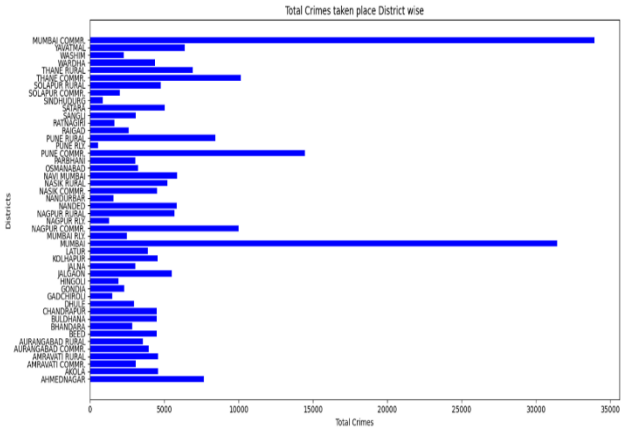


Fig 3: Plot of total crimes taken place District wise.

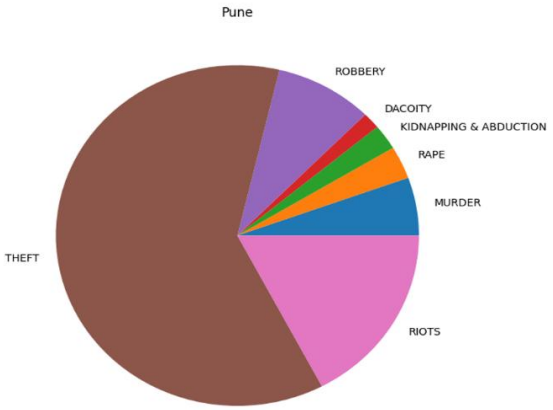


Fig 5: Plot of percentage of different crime taken place in Pune district.

4.1.3 K-Means Clustering Analysis

K-Means clustering is used to segregate the cities into various clusters based on the crime rate occurrences.

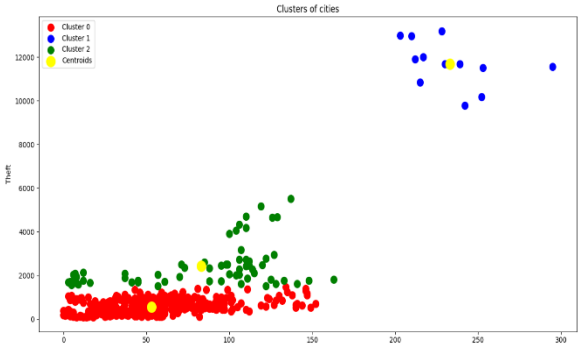
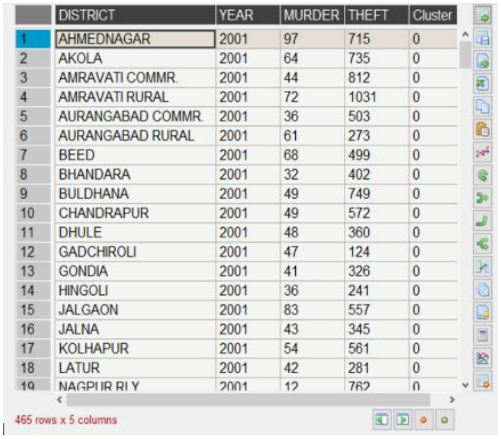


Fig 6: Plot of clusters by using k-means clustering.

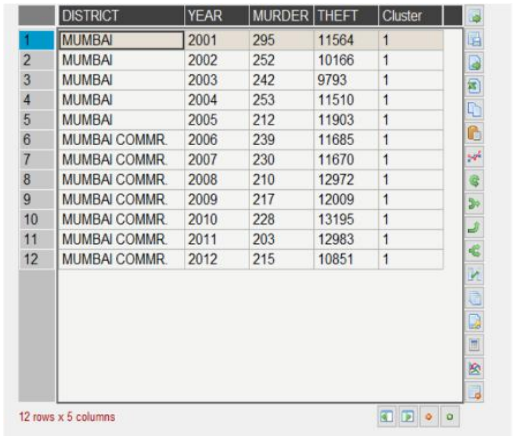
Figure 6 shows that the cities can be segregated into three clusters based on the crime rate occurrences in these cities.



	DISTRICT	YEAR	MURDER	THEFT	Cluster
1	AHMEDNAGAR	2001	97	715	0
2	AKOLA	2001	64	735	0
3	AMRAVATI COMMR.	2001	44	812	0
4	AMRAVATI RURAL	2001	72	1031	0
5	AURANGABAD COMMR.	2001	36	503	0
6	AURANGABAD RURAL	2001	61	273	0
7	BEED	2001	68	499	0
8	BHANDARA	2001	32	402	0
9	BULDHANA	2001	49	749	0
10	CHANDRAPUR	2001	49	572	0
11	DHULE	2001	48	360	0
12	GADCHIROLI	2001	47	124	0
13	GONDIA	2001	41	326	0
14	HINGOLI	2001	36	241	0
15	JALGAON	2001	83	557	0
16	JALNA	2001	43	345	0
17	KOLHAPUR	2001	54	561	0
18	LATUR	2001	42	281	0
19	NAGPUR RLY.	2001	12	762	0

465 rows x 5 columns

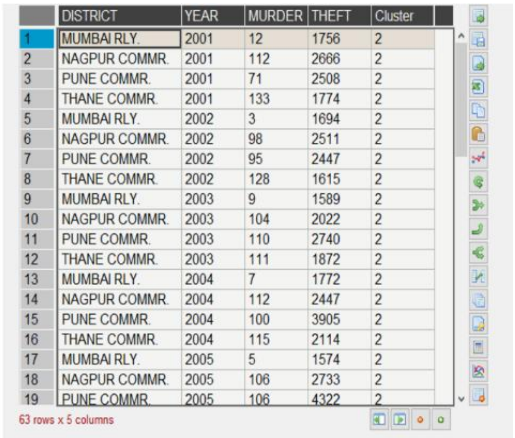
Fig 7: List of cities present in cluster no. 0



	DISTRICT	YEAR	MURDER	THEFT	Cluster
1	MUMBAI	2001	295	11564	1
2	MUMBAI	2002	252	10166	1
3	MUMBAI	2003	242	9793	1
4	MUMBAI	2004	253	11510	1
5	MUMBAI	2005	212	11903	1
6	MUMBAI COMMR.	2006	239	11685	1
7	MUMBAI COMMR.	2007	230	11670	1
8	MUMBAI COMMR.	2008	210	12972	1
9	MUMBAI COMMR.	2009	217	12009	1
10	MUMBAI COMMR.	2010	228	13195	1
11	MUMBAI COMMR.	2011	203	12983	1
12	MUMBAI COMMR.	2012	215	10851	1

12 rows x 5 columns

Fig 8: List of cities present in cluster no. 1



	DISTRICT	YEAR	MURDER	THEFT	Cluster
1	MUMBAI RLY.	2001	12	1756	2
2	NAGPUR COMMR.	2001	112	2666	2
3	PUNE COMMR.	2001	71	2508	2
4	THANE COMMR.	2001	133	1774	2
5	MUMBAI RLY.	2002	3	1694	2
6	NAGPUR COMMR.	2002	98	2511	2
7	PUNE COMMR.	2002	95	2447	2
8	THANE COMMR.	2002	128	1615	2
9	MUMBAI RLY.	2003	9	1589	2
10	NAGPUR COMMR.	2003	104	2022	2
11	PUNE COMMR.	2003	110	2740	2
12	THANE COMMR.	2003	111	1872	2
13	MUMBAI RLY.	2004	7	1772	2
14	NAGPUR COMMR.	2004	112	2447	2
15	PUNE COMMR.	2004	100	3905	2
16	THANE COMMR.	2004	115	2114	2
17	MUMBAI RLY.	2005	5	1574	2
18	NAGPUR COMMR.	2005	106	2733	2
19	PUNE COMMR.	2005	106	4322	2

63 rows x 5 columns

Fig 9: List of cities present in cluster no. 2

Figure 7, figure 8, and figure 9 shows the list of all the cities present in cluster no.0 (i.e., red colour), cluster no.1 (i.e., blue colour), and cluster no.2 (i.e., green colour) respectively.Here red cluster indicate low crime rate, green cluster indicatesa moderate crime rate, and the blue cluster indicatesa high crime rate.Hence from the clustering analysis, we can easily identify the cities having a high risk of criminal activities and take necessary action.

4.1.4 Crime Prediction using Regression Model

In the proposed system the dataset is divided into training and test set. The test set is used for prediction purposes. The random forest algorithm of machine learning is applied to the dataset.

Fig 10: Predicted crime in Pune city in the year 2025.

Using the Regression Model, we can identify the crime type in the future based on the previously occurred crime rates.

5. FUTURE SCOPE

The proposed work mainly focuses on the crime rates around a specific state. work can be extended in the future by using different data mining techniques.

6. CONCLUSION

In this system, the crime is analysed by implementing a k-means clustering algorithm on the crime dataset and future crime type in different cities is predicted using the Random Forest Regression Model. It will help the government officials easily control the crime that occurs in a particular city.

REFERENCES

[1] S.G. Krishnendu, P.P.Lakshmi, L Nitha,Department of Computer Science and IT, Amrita School of Arts and Sciences,Kochi Amrita Vishwa Vidhyapeetham, India, Crime Analysis and Prediction using Optimized K-Means Algorithm(2010)

[2] Gouri Jha, Laxmi Ahuja, Ajay Rana, Criminal Behaviour Analysis and Segmentation using K-means Clustering, 2020 8th International Conference on Reliability, Infocom Technologies and Optimization(ICRITO), Amity University, Noida, 2020

[3] Khushabu A. Bokde, Tiksha P. Kakade, Dnyaneshwari S.Tumsare, Chetan G. Wadhai (April 2018), Department of CSE, Ballarpur Institute of Technology, Crime Analysis Using K-Means Clustering, International Journal of Engineering Research & Technology (IJERT)

[4] Yadav, S., Timbadia, M., Yadav, A., Vishwakarma, R., & Yadav, N. (2017, April). Crime pattern detection, analysis & prediction. In Electronics, Communication,and Aerospace Technology (ICECA), 2017 International Conference of (Vol. 1, pp. 225-230). IEEE

[5] ShijuSathyadevan M.S, Surya Gangadharan: Crime Analysis and Prediction Using Data Mining, in NetworksSoft Computing (ICNSC),(2014) First International Conference.

[6] JesiaQuaderYuki, Md.MahfilQuaderSakib, ZaishaZamal, Khan Mohammad Habibullah, Amit Kumar Das:Predicting Crime Using Time and Location Data(2019).

[7] Malathi. A &Dr. S. SanthoshBaboo: An Enhanced Algorithm to Predict a Future Crime using Data Mining. International Journal of Computer

[8] Malathi, A; Santhosh Baboo, S, 2011 An Enhanced Algorithm

to Predict a Future Crime using Data Mining

[9] ChhayaChauhan1, SmritiSehgal2: A REVIEW: Crime Analysis Using Data Processing Techniques And Algorithms. ICCCA2017)