



CRIME BASED CLUSTERING AND ZONING

Vedant Patil¹, Aniket Desale², Yash Palekar³, Tanishka Patil⁴, Prof. M. J. Patil⁵

Student, Department of Computer Engineering, K. K. Wagh Institute of Engineering Education and Research,
Nashik¹⁻⁴

Professor, Department of Computer Engineering, K. K. Wagh Institute of Engineering Education and Research, Nashik⁵

Abstract: Crime is one of the most predominant and alarming aspects of our society and its prevention is a vital task. Crime analysis is a systematic way of detecting and investigating patterns and trends in crime. Thus, it seems necessary to study the reasons, factors and relations between the occurrence of different crimes and finding the most appropriate ways to control and avoid more crimes. Our System focuses on finding spatial and temporal criminal hotspots. We will make cluster analysis by using the k-means cluster algorithm on the criminal dataset of India. The cluster input is used to create a custom India map with the cluster zones of states. To cluster the crime activities based on some predefined cases and the results of these clustering are compared to find the best suitable clustering algorithm for crime detection. Our System aims to raise people's awareness regarding the dangerous locations and to help agencies to predict future crimes in a specific location within a particular time [1].

Keywords: Clustering, Crime Analysis, Data mining, Hot Spot detection

I. INTRODUCTION

Crime in India has always been on a steep rise and rates of crime occurrence have been increasing by 2 percent each year. With the advancement in technology, crime styles have also evolved and thus become more difficult to catch. The proposed system is aimed to help the law enforcement agencies to implement appropriate measures depending on the type of crime in a certain area. It will assist in predicting the crimes in a certain area by taking into account the data of previous crimes. Today, the collection and analysis of crime-related data are imperative to security agencies. The use of a coherent method to classify these data based on the rate and location of occurrence, detection of the hidden pattern among the committed crimes at different times, and prediction of their future relationship are the most important aspects that have to be addressed. Our faces might disclose more than what we expect, such as race, gender, age, health, emotion, psychology and profession.

II. METHODOLOGY

- In the data pre-processing step, the attributes in the dataset are examined by the frequency analysis method, the insignificant attributes are reduced and the nominal values are converted into numerical values for program optimization.
- The data will then be clustered according to the type and place of the crime occurrence
- Custom map of India will be created which displays these cluster zones in different color-coded sectors for better and easier understanding.
- Image processing will be performed as the image of a suspect will be the input in the system. The system will analyze and match it with the image associated with the crime and inform if the suspect was present at the given time at the given location.

III. SOFTWARE REQUIREMENT SPECIFICATION

A. FUNCTIONAL REQUIREMENTS

- Authorized user should log in to start the application
- Input dataset in the format of .csv or excel sheet
- Input image in jpg or png format with clear distinction of the intended suspect.

B. NON-FUNCTIONAL REQUIREMENTS

Performance requirements

- System will result better if it has proper training and proper thresholding is given

Security requirements

- System will be using login authentication methods for data privacy



C. CONSTRAINTS

Hardware Constraints

- The system should meet minimum hardware specifications as stated in section 4.4

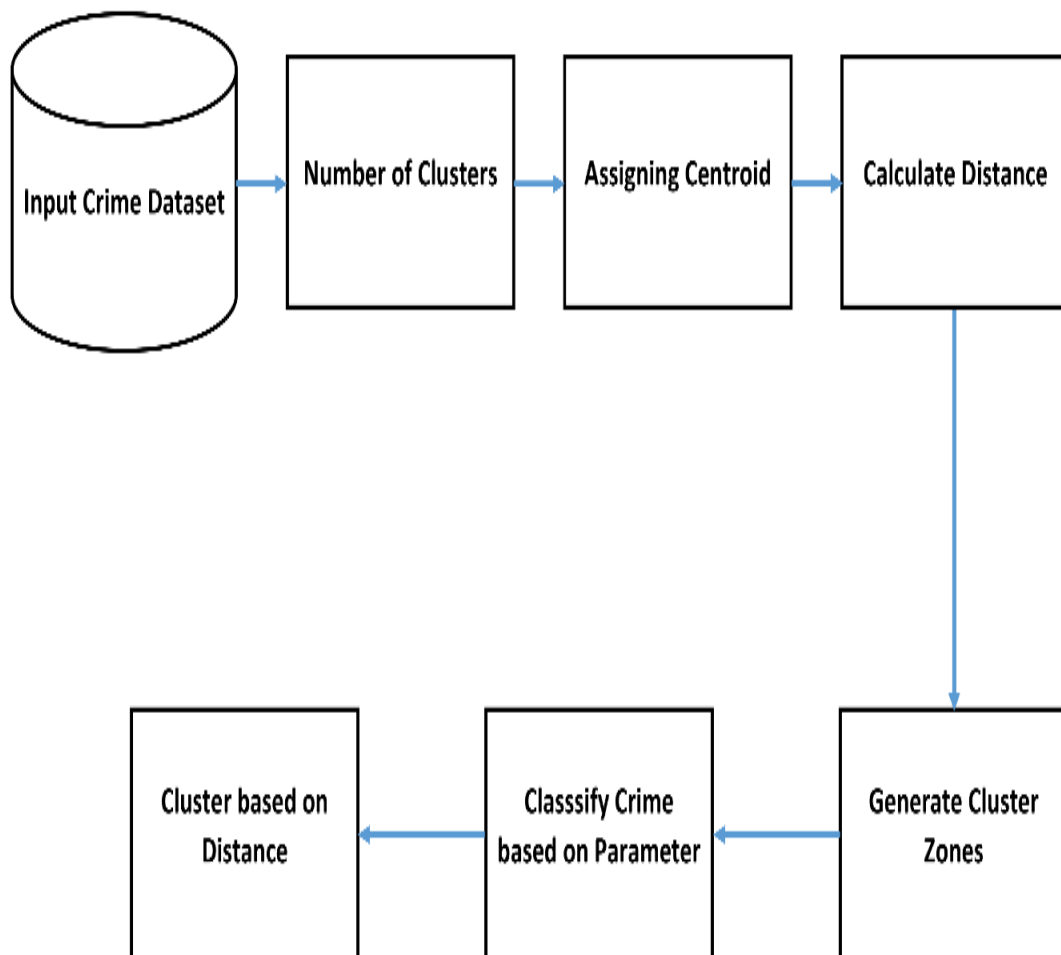
Software Constraints

- All libraries/modules such as scikit-learn, pandas, scipy, and scikit-image should be updated so that system functions properly.

Operational Constraints

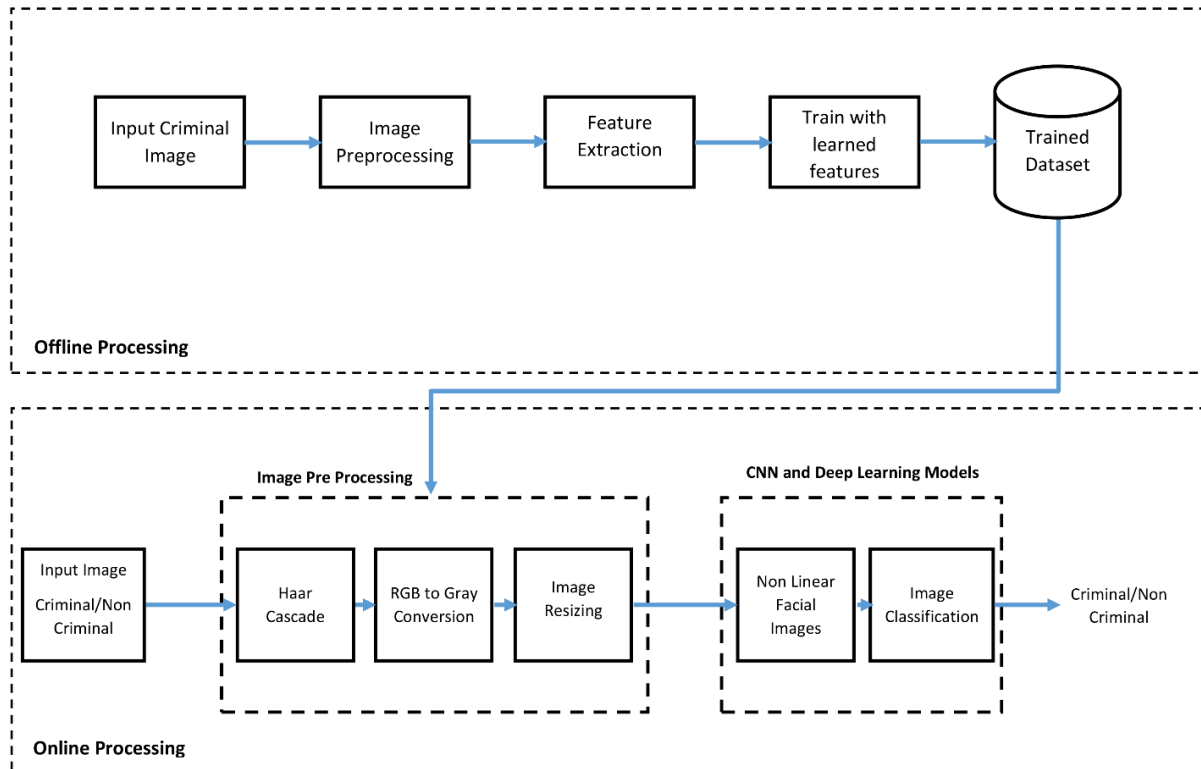
- Working Internet Connection with good speed.

IV. DETAILED DESIGN



Block Diagram

- Input Crime Dataset : Dataset from the repository is uploaded
- Number of Clusters : Clusters are created with respective to crime numbers
- Assigning Centroid : The system assigns centroids for each cluster
- Calculate Distance : The euclidian distance between the centroids and points is calculated
- Generate Cluster Zones : Cluster zones are generated using this distance
- Classify Crime based on Parameter : The crime type is assigned based on the parameter
- Cluster based on Distance : Final Clusters are created



Offline processing

- **Input Image** : An image of a person is uploaded for further analysis
- **Image Preprocessing** : The image is formatted and made suitable for the system
- **Feature Extraction** : It is a dimensionality reduction process, in which, an initial set of the raw data is divided and reduced into more manageable groups
- **Train with Learned Features** : The system is trained with these features
- **Trained Dataset** : The system is successfully trained

Online processing

- **Input Criminal Image** : An image of a criminal is uploaded for matching
- **Haar Cascade** : This feature-based object detection classifier is used to differentiate the features of the image such as face, mouth, eyes, etc
- **RGB to Gray Conversion** : The RGB values are converted to grayscale using the NTSC (National Television System Committee) formula: $0.299 \text{ Red} + 0.587 \text{ Green} + 0.114 \text{ Blue}$. This formula closely represents the average person's relative perception of the brightness of red, green, and blue light
- **Image Resizing** : Resizing of the image is done
- **Non-Linear Facial Images** : Non-Linear images are converted to linear images for better analysis
- **Image Classification** : It is the process of categorizing and labeling groups of pixels or vectors within an image based on specific rules
- **Result** : Criminal/Non-criminal

V. IMPLEMENTATION

OVERVIEW OF PROJECT MODULES

- **Input Crime Dataset** : Dataset acquired from NCRB repository comes with a predefined order of districts according to states. This dataset will be used as an input to the system for zoning.
- **Cluster Algorithm** : K means algorithm is used to cluster the districts according to the predominant crime type. This will help us create a group of districts with a similar crime rate of a particular type of crime.
- **Graphical Representation of Crime Data** : The clustered data is displayed in various ways. To get a comparison of districts regarding crime, a bar graph is shown with the drop-down menu to choose districts. Similarly, crimes can be



compared with each other with respect to the districts. External website: paintmaps.com is used to generate a digital/graphical representation of the data. Users can hover over the district to get figures of the crime chosen.

- Haar Cascade : This algorithm is used to detect the face of a criminal if present in the input photo. It is then matched with the criminal dataset already fitted in the system.

A. EFFICIENCY ISSUES

- Time required to acquire dataset containing proper data regarding crime sections.
- Proper dataset would determine and affect the efficiency of the model.
- Time is taken to organize the cluster zones based on the crime category on the map graphically.

VI. CONCLUSION

Crimes in India are rising at an alarming rate because of the factors such as the increase in poverty, migration, unemployment, frustration, illiteracy and corruption. Crime investigating agencies search the database of criminals manually or with some computer data analyst which is a tedious process and takes much more time. So, to contribute toward combating crimes and identifying criminals, we proposed an integrated technology of CDCI (Crime Detection and Crime Identification) using DMT (Data Mining Techniques) for Indian Districts. The system is able to identify and display crime clusters in an easy-to-understand and comprehending way for better implementation of law enforcement and will also help in the detection of a specific individual at the crime scene.

This system is designed especially for the general public to be up-to-date about their surrounding crime situations and take corresponding measures. People can check their city's data and compare it to others. This system can be further developed into a mobile application for more ease of use. People can easily access data and also be aware of the areas. People can upload images of suspicious people to get a check of criminal/non-criminal person. This system can also be used in law enforcement agencies like the police force. It can determine which areas need what kind of law enforcement. This will help control the crime rates and avoid harm to the public.

ACKNOWLEDGMENT

It gives us great pleasure in presenting the preliminary report on 'CRIME BASED CLUSTERING AND ZONING'. We would like to take this opportunity to thank our internal guide, **Prof. Megha Patil**, for giving us all the help, guidance and motivation we needed from time to time. We are really grateful to her for her kind support. Her valuable suggestions were very helpful.

We are also grateful to **Prof. Dr. S. S. Sane**, Head of Computer Engineering Department, K. K. Wagh Institute of Engineering Education and Research, Nashik for his indispensable support, and suggestions.

In the end, our special thanks to our internal guide again for providing various resources such as a laboratory with all needed software platforms, and a continuous Internet connection, for our Project.

REFERENCES

- [1]. L. S. Thota, M. Alalyan, A. -O. A. Khalid, F. Fathima, S. B. Changalasetty and M. Shiblee, "Cluster based zoning of crime info," 2017 2nd International Conference on Anti-Cyber Crimes (ICACC), 2017, pp. 87-92, doi: 10.1109/Anti-Cybercrime.2017.7905269. J. Breckling, Ed., The Analysis of Directional Time Series: Applications to Wind Speed and Direction, ser. Lecture Notes in Statistics. Berlin, Germany: Springer, 1989, vol. 61
- [2]. Sunil Yadav, Meet Timbadia, Ajit Yadav, Rohit Vishwakarma, and Nikhilesh Yadav "Crime pattern detection, analysis, prediction" In 2017 International conference of Electronics, Communication and Aerospace Technology (ICECA), vol. 1, pp. 225-230. IEEE, 2017
- [3]. Devendra Kumar Tayal, Arti Jain, Surbhi Arora, Surbhi Agarwal, Tushar Gupta, and Nikhil Tyagi "Crime detection and criminal identification in India using data mining techniques" AI society, vol. 30, no. 1, 2015 (2002) The IEEE website. [Online]. Available: <http://www.ieee.org/>
- [4]. Mehmet Sevri, Hacer Karacan, and M. Ali Akcayol "Crime analysis based on association rules using apriori algorithm" International Journal of Information and Electronics Engineering, Vol. 7, no. 3 pp. 99, 2017 FLEXChip Signal Processor (MC68175/D), Motorola, 1996.
- [5]. "Using Data Mining Techniques to Analyze Crime patterns", the Libyan National Crime Data Zakaria Suliman Zubi et al.
- [6]. Sang C. Sug, "Practical Applications of Data Mining" Jones ; Bartlett, 2012
- [7]. T. Gonzalez, "On the computational complexity of clustering and related problems," R. Drenick and F. Kozin, Eds.
- [8]. D. Arthur and S. Vassilvitskii, "k-means++: the advantages of careful seeding"