

A PRELIMINARY PROJECT REPORT ON
CRIME BASED CLUSTERING AND ZONING

SUBMITTED TO SAVITRIBAI PHULE PUNE UNIVERSITY, PUNE IN
PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE AWARD OF
THE DEGREE OF

**BACHELOR OF ENGINEERING
(COMPUTER ENGINEERING)**

SUBMITTED BY

Aniket Desale	Exam No:B150134228
Vedant Patil	Exam No:B150134301
Yash Palekar	Exam No:B150134285
Tanishka Patil	Exam No:B150134299

**UNDER THE GUIDANCE OF
PROF M.J.PATIL**



**Department of Computer Engineering
K. K. Wagh Institute Of Engineering Education & Research
Hirabai Haridas Vidyanagari, Amruttadham, Panchavati,
Nashik-422003
Savitribai Phule Pune University
A. Y. 2021-2022**

Acknowledgment

It gives us great pleasure in presenting the preliminary project 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, suggestions.

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

Aniket Desale

Vedant Patil

Yash Palekar

Tanishka Patil

(B.E. Computer Engg.)

Abstract

Crime is one of the most predominant and alarming aspects in 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 reasons, factors and relations between 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 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. 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

INDEX

1	Introduction	1
1.1	Project Idea	2
1.2	Motivation of the Project	2
1.3	Literature Survey	2
2	Problem Definition and scope	6
2.1	Problem Statement	7
2.1.1	Goals and objectives	7
2.1.2	Assumption and Scope	7
2.2	Methodology	8
2.3	Outcome	8
2.4	Type of Project	8
3	Project Plan	9
3.1	Project Timeline	10
3.2	Team Organization	10
3.2.1	Team structure	10
4	Software requirement specification	11
4.1	Functional Requirements	12
4.2	Non Functional Requirements	12
4.3	Constraints	12
4.4	Hardware Requirements	12
4.5	Software Requirements	13

5 Detailed Design	14
5.1 Architectural Design(Block Diagram)	15
5.1.1 Block Diagram	15
5.1.2 Offline processing	16
5.1.3 Online processing	16
5.2 Algorithm Used	18
5.3 Data design	18
5.3.1 Data Structure	18
5.4 Component design/ Data Model	19
6 Project Implementation	20
6.1 Overview of Project Modules	21
6.2 Tools and Technology used	21
7 Results and Discussion	22
7.1 Data Set	23
7.2 Performance Parameters	23
7.3 Efficiency Issues	23
7.4 Confusion Matrix	24
7.5 Software Testing	24
7.5.1 Test Cases	24
7.5.2 Test Results	28
8 Conclusion and Future Work	34
8.1 Conclusion	35
8.2 Future Work	35
Annexure A Plagiarism Report	38

List of Figures

3.1	Timeline Chart	10
5.1	Block Diagram of Clustering	15
5.2	Training and Testing model.	16
5.3	Class Diagram	19
5.4	Component Diagram	19
7.1	Precision Recall Formulas	23
7.2	Confusion Matrix	24
7.3	Input Dataset	28
7.4	Clustered Dataset	28
7.5	Bar Graph Representation	29
7.6	Mapping of Dataset	29
7.7	Sketch based matching	30
7.8	Sketch generation	30
7.9	Sketch matching	31
7.10	Image matching	31
7.11	Feature matching	32
7.12	Key descriptor points	32
7.13	Image Matching results	33

List of Tables

1.1	Literature Review	5
7.1	GUI Testing Testcases	24
7.2	Zoning Testing Testcases	25
7.3	Image Matching Testing Testcases	26
7.4	Sketch based Testing Testcases	27

CHAPTER 1

INTRODUCTION

1.1 PROJECT IDEA

- Crime in India has always been at a steep rise and rates of crime occurrence has been increasing by 2 percent each year. With 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.

1.2 MOTIVATION OF THE PROJECT

- Today, 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.

1.3 LITERATURE SURVEY

- **Crime pattern Detection, Analysis prediction Sunil Yadav et al. [2]** developed a model for crime prediction to improve the accuracy. In this model data set are collected from police department that contain, the number of individuals arrested and number of individual crimes committed with various other attributes. WEKA preprocessing techniques were applied to clean the unnecessary data. K-means algorithm was used to cluster the data such as high number of individuals committed in crime and, low number of individuals committed in crime. Apriori algorithm (association mining) was used to discover the frequent item set. In this model Apriori Algorithm is applied in clustered data and that data is divided into high and low values. The outcome of Apriori demonstrated a relationship among the persons captured during year

and person released in the same year. In this model Naïve Bayes algorithm was used to classify data like, number of crimes committed by a specific age group. In this model linier/linear regression is applied to predict the number of persons released from the crime against the number of hearings finalized throughout the year. But this paper fails to discuss the accuracy metrics of the system.

- **Crime detection and criminal identification in India using data mining techniques Devendra Kumar Tayal et. al.** [3] proposed the intergraded system of crime detection and criminal identification by applying data mining techniques for Indian cites. In this system data are collected from NCRB (National Crime Record Bureau), CPJ (Committee to protect journalists), and other Web source, from the period of 2002- 2012. Data pre-processing techniques are applied to clean and convert the data into a structured format. In this model seven Indian cities were selected namely, Bengaluru, Delhi, Hyderabad, Kolkata, Mumbai, Jaipur and Pune and K-means clustering algorithm was applied to cluster the cities based on crime rate. This model still needs enhancements on data collection, the accuracy of crime classification and other security aspects.
- **Crime Analysis Based on Association Rules Using Apriori Algorithm Mehmet Sevri et al.** [4] displayed an itemized Crime Analysis that depended on data mining Association Rules. The author applied the Apriori Algorithm to make an association rule to remove the connection among present and past Crime violations. In this framework data collection was done from NIBRS (National Incident-Based Reporting System) crime repository and criminal records from USA. Dataset contained 48 types of crime taken from nearly 5 million crime stories. The crime dataset contained various attributes namely: State, Population Group, Incident Date and Hour, Location type, weapon / force, Type property loss, property description, sex of victim, Relationship of victim to offender, offender age and sex. The proposed framework arranged the dataset

into five phases namely Data pre-processing, data encoding, creating transactions in the dataset, creating frequent item sets which provide minimum support value and creating association rules which provide minimum confidence value out of the created item sets. In data pre-processing phase, the dataset attributes are examined using frequency analysis method to reduce irrelevant attributes and nominal values are converted into numerical values for program optimization. Data preprocessing outcome results are then coded in data encoding phase. In transaction creation phase, Apriori algorithm was applied to create the transactions dataset, followed by identifying frequent itemset that provides the minimum support value, and in the final phase association rules were framed to provide the minimum confidence value from the frequent item sets. This overall framework covers a detailed analysis of crimes. In this system, 5 million reports were analyzed but the system could only identify the recent 300 incidents. Other missing outliers data is not discussed.

- **Using Data Mining Techniques to Analyze Crime patterns in the Libyan National Crime Data** Zakaria Suliman Zubi et al. [5] Proposed a framework for crime and criminal analysis. In this system, dataset was collected from the Libyan national crime data. Dataset contained more than 350 crimes and criminal records. WEKA mining software Numeric- To-Nominal function was applied to pre-process the collected data. In this framework K-Means clustering algorithm was applied to group the crimes and criminals. It additionally gave the general statistical factual information about Criminal age versus Crime types. In this framework Apriori Algorithm was used for Association rule mining to identify frequent crime rate. Finally, both the algorithms displayed a promising outcome. The model is solely based for the Lybian government and would not function for other countries. Also the model does not include additional features like image processing and displaying custom map of India.

Paper title	Algorithm Used	Merits	Demerits
Crime pattern Detection, Analysis and prediction	Apriori algorithm	The model predicated; Each 10 rape case trials were finished and only 2.5 to 3 people are sentenced for the rape charge.	Accuracy metrics are not deeply discussed
Crime detection and criminal identification in India using data mining techniques	K-means clustering algorithm	This model predicted; Delhi has the highest number of rape rate between other Indian cities. These helped the cities in providing tight security for women.	Security measures and accuracy is not as expected
Crime Analysis Based on Association Rules Using Apriori Algorithm	Apriori algorithm	The framework identified 300 frequent item sets and 368 relationships. This solid system applied Association rules to find the Minimum Support Value whose value was 0.05 and the minimum confidence value was 0.06	It analysed 5 million reports but could only identify 300 recent itemsets
Using Data Mining Techniques to Analyze Crime patterns in the Libyan National Crime Data	K-means clustering algorithm, Apriori algorithm	This system identified 32,33,34,36 aged group criminals committed the high rates crime and 15, 16, 43,44,51,52 aged group criminals committed the low rates	The model is solely based for the Lybian government and would not function for other countries. Also the model does not include additional features like image processing and displaying custom map of India.

Table 1.1: Literature Review

CHAPTER 2

PROBLEM DEFINITION AND SCOPE

2.1 PROBLEM STATEMENT

- To study reasons, factors and relations between occurrence of different crimes and finding the most appropriate ways to control and avoid more crimes.
- To identify criminal pattern to pace up the mechanism of crime solutions.
- To learn and identify a person using his facial images

2.1.1 Goals and objectives

Goal and Objectives:

- To classify clustered crimes based on frequency of crime occurrence during different years.
- To find the difference between the facial characteristic of criminal and non-criminal faces.
- To execute crime analysis from huge criminal data, an appropriate scientific field needs to be chosen.

2.1.2 Assumption and Scope

- Data mining is used extensively in terms of analysis, investigation and discovery of patterns for occurrence of different crimes.
- We propose a theoretical model based on data mining techniques such as clustering and classification to dataset

2.2 METHODOLOGY

- In the data pre-processing step, the attributes in the dataset are examined by 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 type and place of the crime occurrence
- Custom map of India will be created which displays these cluster zones in different colour coded sectors for better and easier understanding.
- Image processing will be performed as image of a suspect will be the input in the system. System will analyse 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

2.3 OUTCOME

- To understand the concept of data mining and machine learning which can be used for finding criminal patterns and behaviors.

2.4 TYPE OF PROJECT

- The project Crime Based Clustering and Zoning is application oriented.
- The Crime Based Clustering and Zoning comes under the domain of Data Mining and Image processing.

CHAPTER 3

PROJECT PLAN

3.1 PROJECT TIMELINE

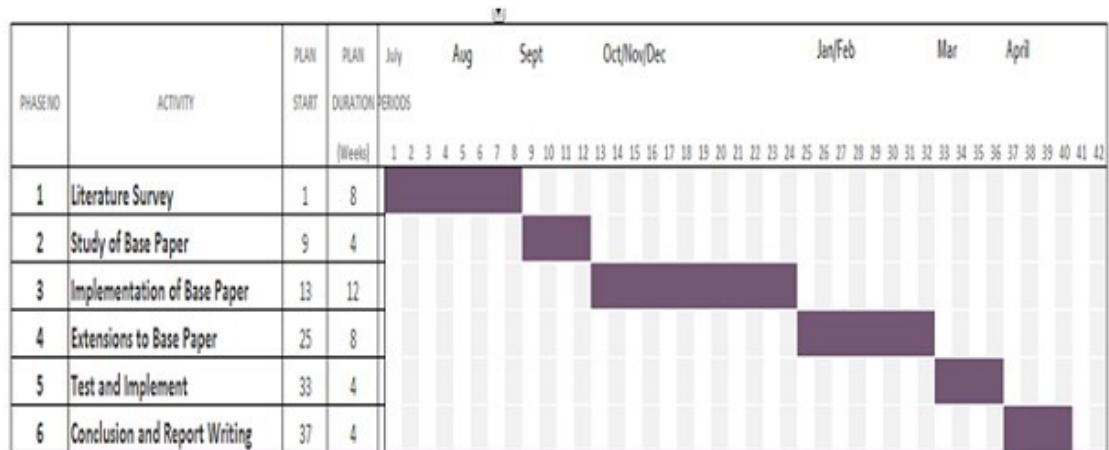


Figure 3.1: Timeline Chart

3.2 TEAM ORGANIZATION

3.2.1 Team structure

- Aniket Desale-
Development Of Software as per requirement.
- Vedant Patil-
Finding improvements to existing methodology (from research papers, articles etc), Designing the system flow and input pre-processing implementation
- Yash Palekar-
Implementation and procurement of available data set.
- Tanishka Patil-
Implementation of image processing algorithm to the application.

CHAPTER 4

SOFTWARE REQUIREMENT

SPECIFICATION

4.1 FUNCTIONAL REQUIREMENTS

- Authorized user should login 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 intended suspect

4.2 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.

4.3 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, scikit-image should be updated so that system functions properly.

Operational Constraints

- Working Internet Connection with good speed.

4.4 HARDWARE REQUIREMENTS

- Hard Disk 40GB and above
- RAM 1 GB and above
- Processor P4 and above

4.5 SOFTWARE REQUIREMENTS

- Software: VS2010
- Tools: .Net Framework Tools
- Dataset: Criminal Dataset

CHAPTER 5

DETAILED DESIGN

5.1 ARCHITECTURAL DESIGN(BLOCK DIAGRAM)

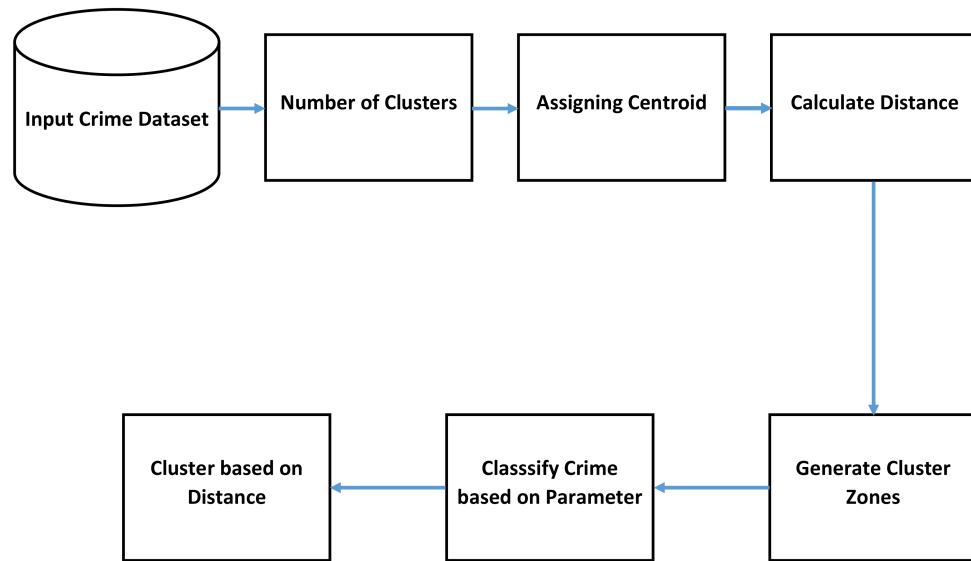


Figure 5.1: Block Diagram of Clustering

5.1.1 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 parameter
- Cluster based on Distance : Final Clusters are created

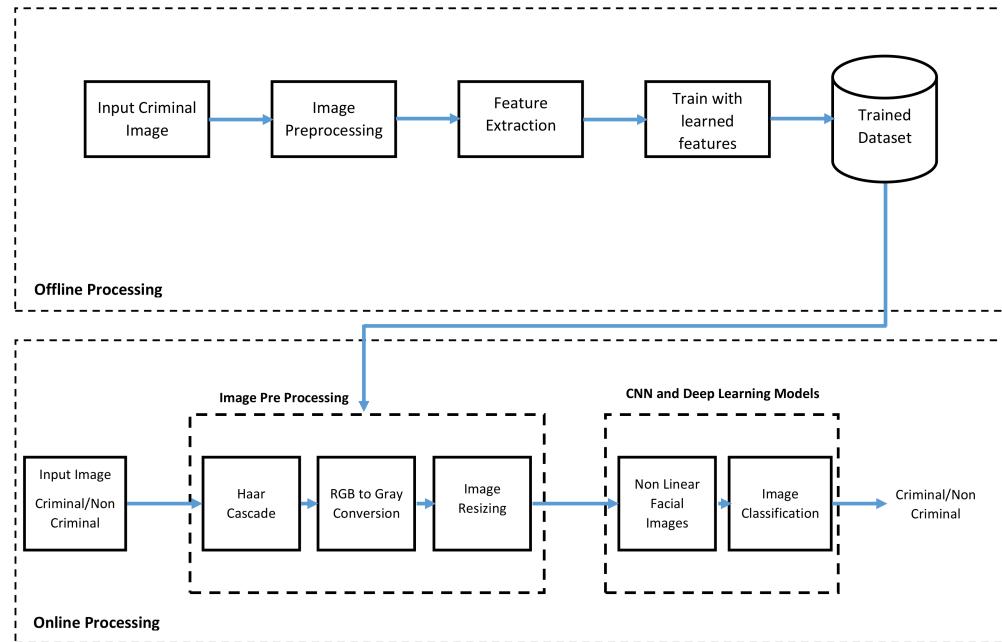


Figure 5.2: Training and Testing model.

5.1.2 Offline processing

- Input Image : An image of a person is uploaded for further analysis
- Image Preprocessing : 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 to more manageable groups
- Train with Learned Features : System is trained with these features
- Trained Dataset : The system is successfully trained

5.1.3 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

5.2 ALGORITHM USED

K-Means Clustering

- The process will follow easy and yet effective method of dividing the whole criminal data sets into a number of clusters based on the type of crime occurred and the place where it happened.
- After defining number of clusters, assign centroids for each and every cluster.
- The next step is to assign every data point to the nearest centroid and we compute the mean of all data points in one cluster and move the centroid to computed distance and form a new centroid for each cluster
- The system will assign each centroid with the respective crimes associated in that vicinity and type of crime

5.3 DATA DESIGN

A description of all data structures including internal, global, and temporary data structures, database design (tables), file formats.

5.3.1 Data Structure

Internal software data structure

- Classification parameters and Class labels are an internal data structure.

Global data structure

- Features and Trained dataset are the global data structure.

Temporary data structure

- Image and attributes features are the temporary data structure.

5.4 COMPONENT DESIGN/ DATA MODEL

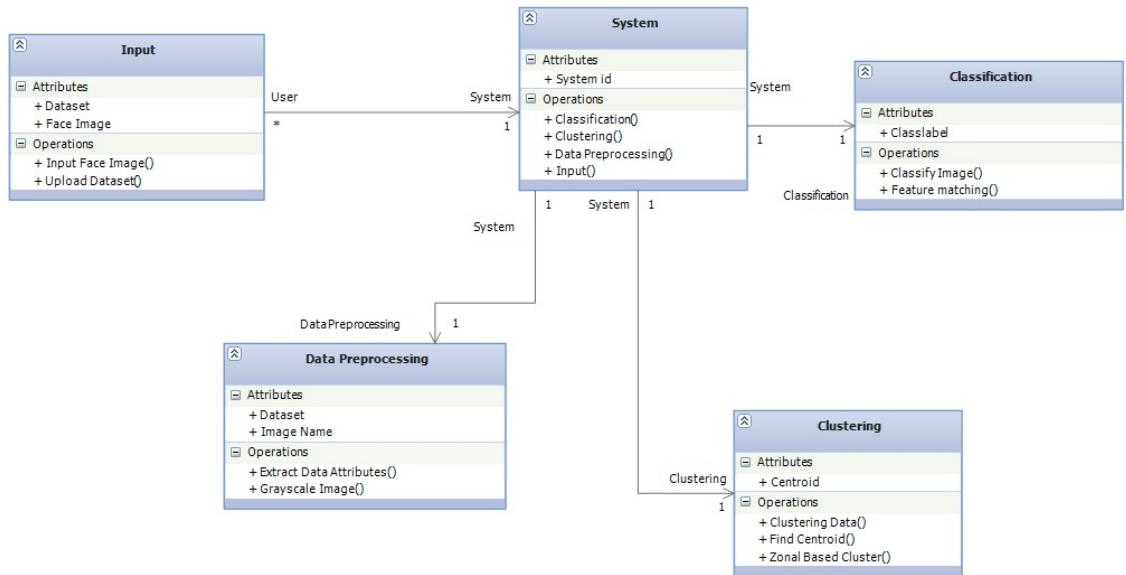


Figure 5.3: Class Diagram

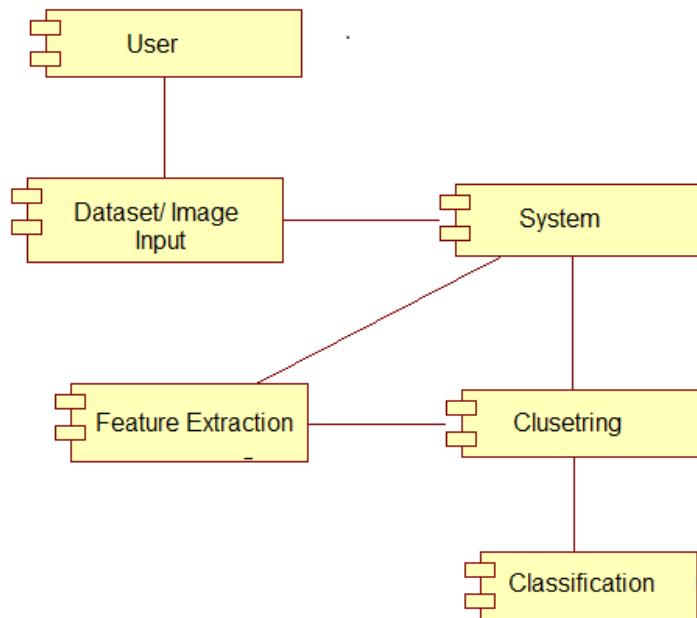


Figure 5.4: Component Diagram

CHAPTER 6

PROJECT IMPLEMENTATION

6.1 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 a 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 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, bar graph is shown with 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 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 criminal if present in the input photo. It is then matched with the criminal dataset already fitted in the system.

6.2 TOOLS AND TECHNOLOGY USED

- Operating System: Windows 7 or more.
- IDE: Visual Studio
- Language: C sharp programming
- Framework: .NET
- Libraries: OpenCV, Emgu CV, AForge

CHAPTER 7

RESULTS AND DISCUSSION

7.1 DATA SET

The analysis of Crime patterns in India is performed on a dataset with the demographic information of crimes in various districts of India. The dataset is taken from National Crime Records Bureau (NCRB). This repository contains crime records of India based on district and type of crime committed. It also contains the total population of the state and total crimes committed with respect to the population.

7.2 PERFORMANCE PARAMETERS

Experiments will be conducted in a form of Precision ,Recall, Accuracy.

$$\text{Precision} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Positive}}$$

$$\text{Recall} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}}$$

Figure 7.1: Precision Recall Formulas

- The precision of our project is 93%.
- The recall is 95%.
- The accuracy is 94%.

7.3 EFFICIENCY ISSUES

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

7.4 CONFUSION MATRIX

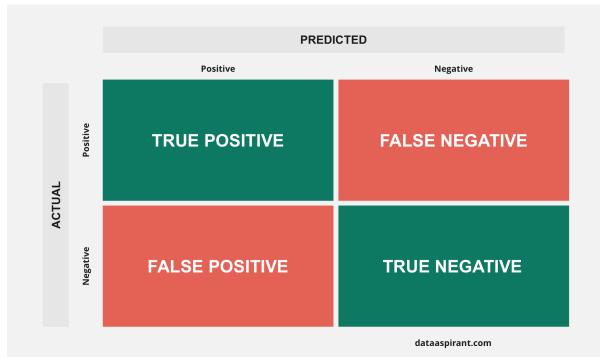


Figure 7.2: Confusion Matrix

7.5 SOFTWARE TESTING

7.5.1 Test Cases

7.5.1.1 GUI Testing

Test Case ID	Test	Description	Result
1	Menu Navigation	User clicks on particular menu to navigate on that page	The respective page gets open
2	Display the uploaded image	The image uploaded by the user should be visible	Image is displayed
3	Result Display	After successfully uploading the image, result should be displayed	A popup is generated displaying the result

Table 7.1: GUI Testing Testcases

7.5.1.2 Zoning Testing

Test Case ID	Test	Description	Result
1	Cluster Dataset	The dataset should be clustered according to crime	The clusters are made and displayed
2	Give dropdown menu	Dropdown menu should be provided to select personalised viewing preference	User is given the option to select districts as well as crime types
3	Year selection	User should be able to select the year as per preference	A dropdown menu is used to give options for year selection
4	Bar graph display	Bar graph should be displayed comprising different features in the dataset	Two types of bargraph are displayed according to preference
5	Mapping	The selected data should be displayed in a color coded map	The application directs the user to a website where a graphical map is presented with color coding

Table 7.2: Zoning Testing Testcases

7.5.1.3 Image Processing Testing

Test Case ID	Test	Description	Result
1	Input valid image	User uploads valid input	The image is accepted
2	Invalid input	User uploads invalid image type	Error message generated for invalid input
3	Feature extraction	Features should be highlighted	Features like face, mouth, eyes are highlighted
4	Image matching	Matched image should be presented	The matched image is displayed and given details

Table 7.3: Image Matching Testing Testcases

7.5.1.4 Sketch Based Testing

Test Case ID	Test	Description	Result
1	Input valid sketch	User generates valid sketch	The sketch is accepted
2	Personalized sketching	User is free to generate personalized sketch	dropdown options are given to generate sketch
3	Matching sketch	Sketch is matched with the dataset	The matched image is displayed

Table 7.4: Sketch based Testing Testcases

7.5.2 Test Results

Crime Based Clustering and Zoning

Criminal Dataset		Dataset												
		2020	Load Dataset	Refresh										
Preprocess Data	District	Sum of RAPE	Sum of DOWRY DEATHS	Sum of COUNTERFEITING	Sum of CHEATING	Sum of RIOTS	Sum of AUTO THEFT	Sum of BURGLARY	Sum of KIDNAPPING AND ABDUCTION OF WOMEN AND CHILDREN	Sum of THEFT	Sum of ROBES			
Clustering	AHMEDN...	653	91	82	2023	4742	4476	5248	354	14405	1543			
	AKOLA	299	74	107	1023	1513	1506	3351	281	10762	658			
	AJORAVATI	376	99	74	1126	2208	2747	6114	450	2943	867			
	AURANGAB...	115	594	40	2247	453	460	5154	513	117546	14553			
	BED...	322	112	16	925	456*	1699	3577	259	7119	914			
	BHANDARA	338	36	46	437	774	672	2315	121	5252	198			
	BULDHANA	381	320	52	960	2432	1323	2541	223	8596	415			
	CHANDRA...	469	62	38	713	1383	1394	3480	219	8285	379			
	DHULE	178	66	19	777	1678	1118	1675	114	4158	290			
	GADCHIR...	245	14	13	144	317	207	1161	94	1606	123			
	GOGAB...	146	35	44	386	459	653	2274	197	1409	207			
	HINGOLI	175	40	18	596	1393	155	1136	101	2200	391			
	JALGAON	485	115	66	2505	2825	1911	2925	266	8571	566			
	JALNA	193	271	16	892	2110	1152	2254	133	4531	623			
	KOLHAPU...	348	245	47	2470	2521	3125	3294	196	8448	444			
	LATUR	231	387	32	937	2518	1450	2456	300	4627	500			
	MUMBAI	219	175	131	20679	3432	4149	34115	1328	161905	5633			
	NAGPUR	1135	159	329	3863	4504	12530	15899	719	49846	3854			
	NANDED	280	280	40	1660	356	1166	443	348	6291	907			
	NANDURB...	221	17	20	422	1089	485	865	92	2657	126			

Figure 7.3: Input Dataset

Crime Based Clustering and Zoning

Criminal Dataset		Clustering													
Preprocess Data	Input Document	<p>(CLUSTER 1) AJRANGABAO,415,394,40,2247,4333,4645,5154,313,13746,1453,440,334,105,415,851,19,68,75,25 ("19,68,75,25") BHANDARA,338,36,46,437,774,472,2151,12,5526,198,146,28,343,32,147,21,15,79,65 ("21,15,79,65") CHANDRAPUR,409,62,38,703,153,154,340,219,8285,379,262,80,613,469,32,22,28,79,67 ("22,28,79,67") GADCHIROLI,245,14,13,144,317,207,1161,35,1696,123,116,104,694,245,402,20,17,79,98 ("20,17,79,98") GOVINDA,348,35,44,139,495,531,2278,197,4900,210,67,432,348,207,21,49,80,13, ("21,49,80,13") JALGAON,485,115,66,505,2825,1911,2926,8571,566,338,272,1007,495,636,21,57,31, ("21,57,31") KOLHAPUR,340,245,47,2470,2521,313,3284,95,844,444,301,167,719,349,45,16,57,24, ("16,57,24,74,34") LATUR,231,387,31,397,2516,1480,2485,300,4627,500,302,137,181,27,442,18,4,76,5 ("18,4,76,5") OSMANABAO,308,23,4,1079,1990,789,2245,195,4396,344,211,216,547,308,276,18,11,76,62 ("18,11,76,62") SATARA,370,73,58,1453,2114,3026,4140,234,9795,898,327,279,728,369,394,17,69,74,03 ("17,69,74,03") (CLUSTER 2) AHMEDNAGAR,653,31,32,2023,4742,4476,5249,354,14405,1543,514,477,1210,653,15,19,69,75, ("19,69,75")</p>													
Clustering	User Defined Cluster:	2	Clustering	Clear	Time : 0.11	Total Iteration:									
Single Graph															
Overall Graph															
MapZoning															
Cancel															

Figure 7.4: Clustered Dataset

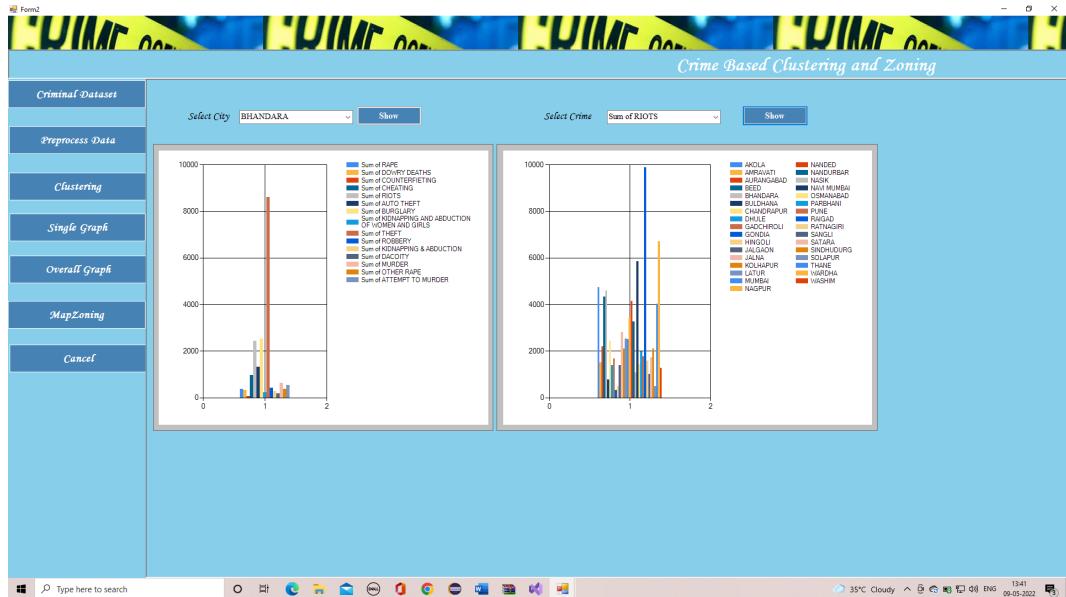


Figure 7.5: Bar Graph Representation

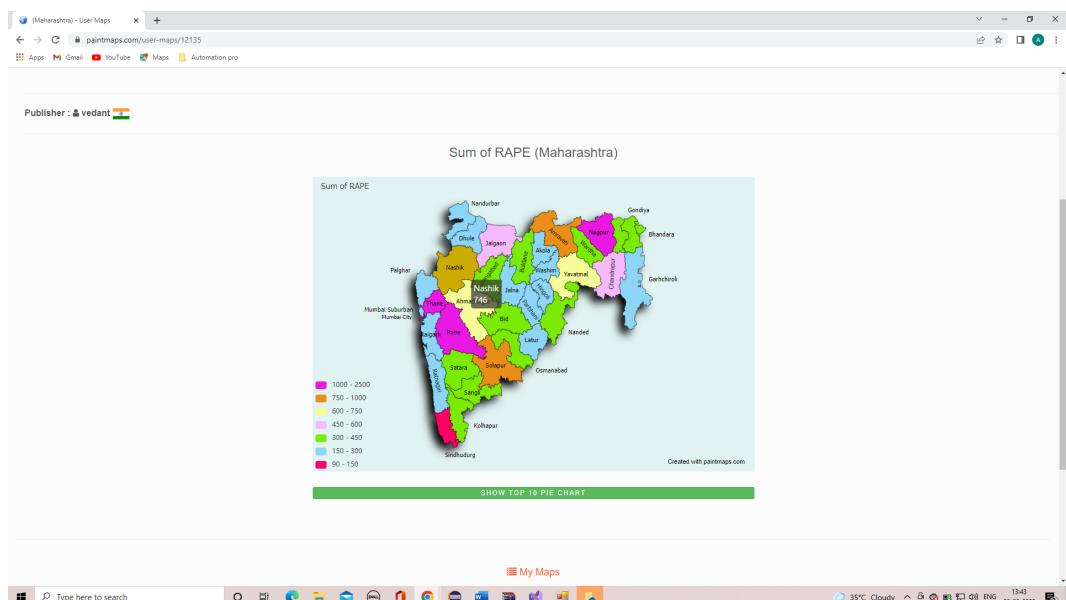


Figure 7.6: Mapping of Dataset

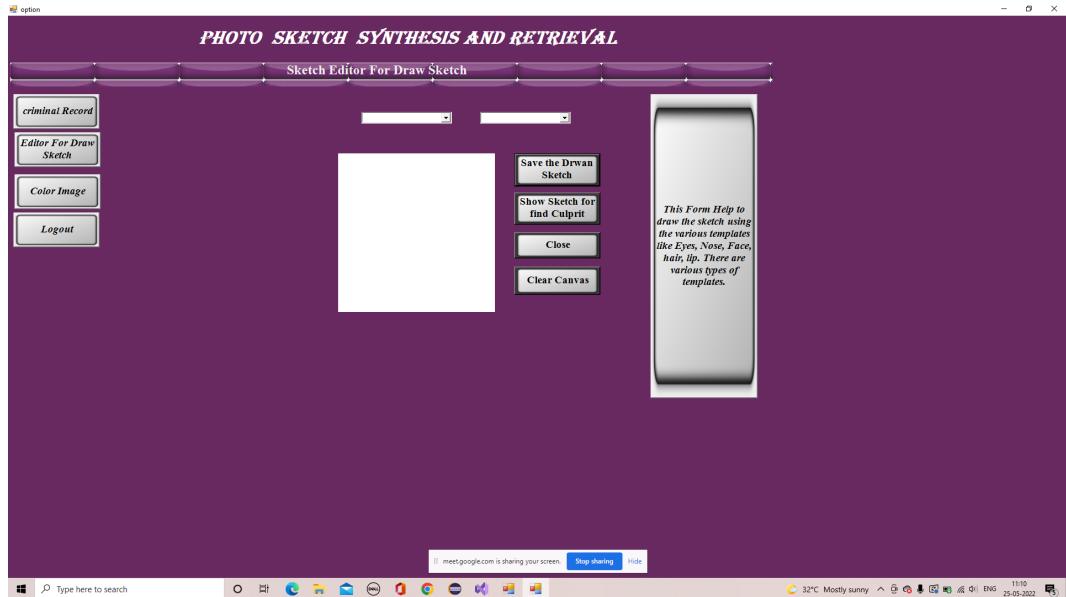


Figure 7.7: Sketch based matching

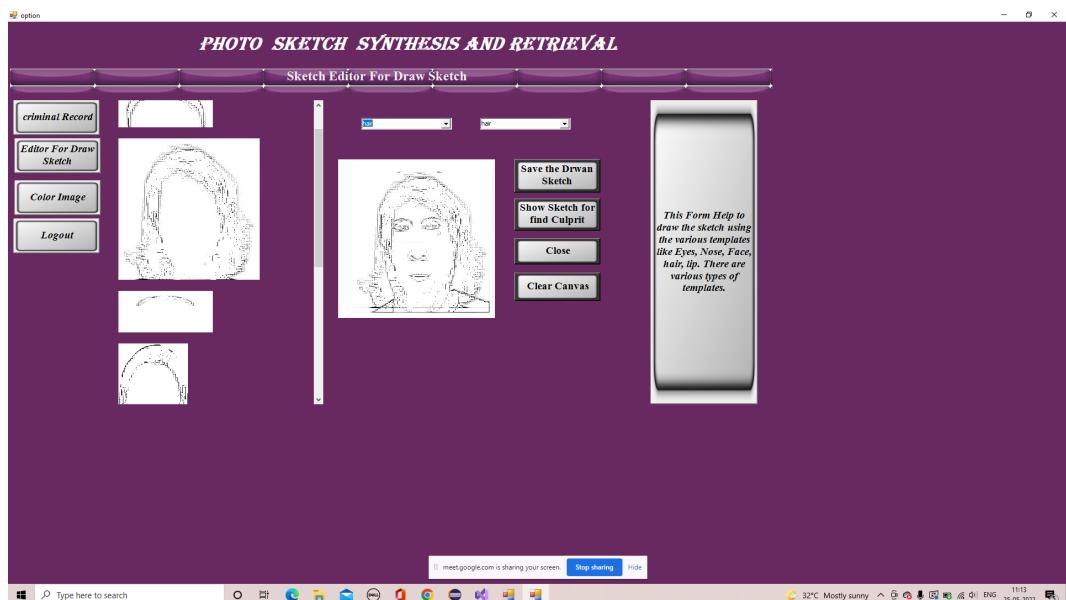


Figure 7.8: Sketch generation

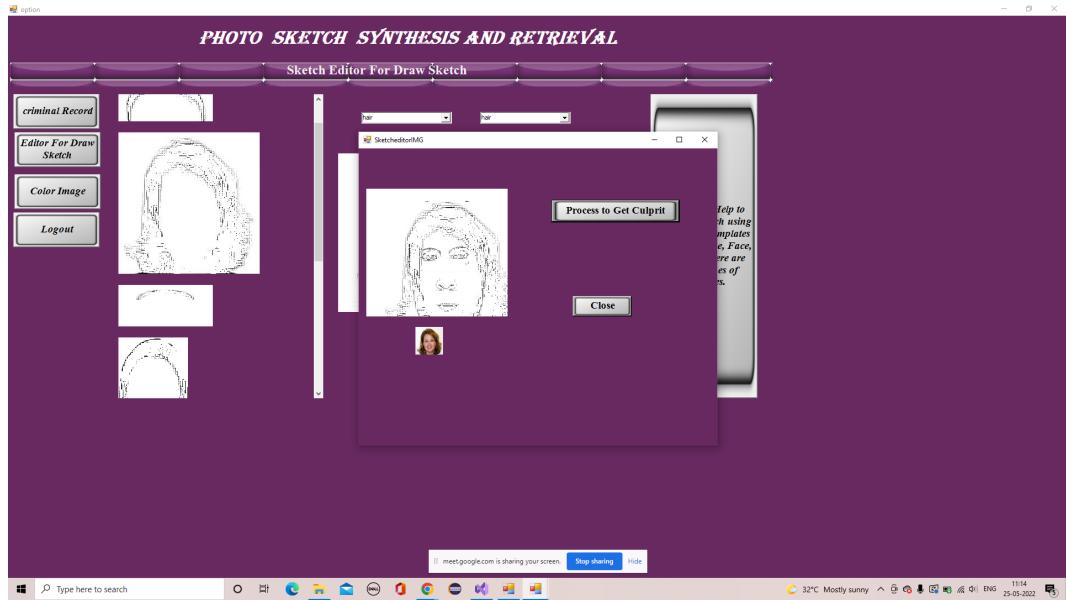


Figure 7.9: Sketch matching

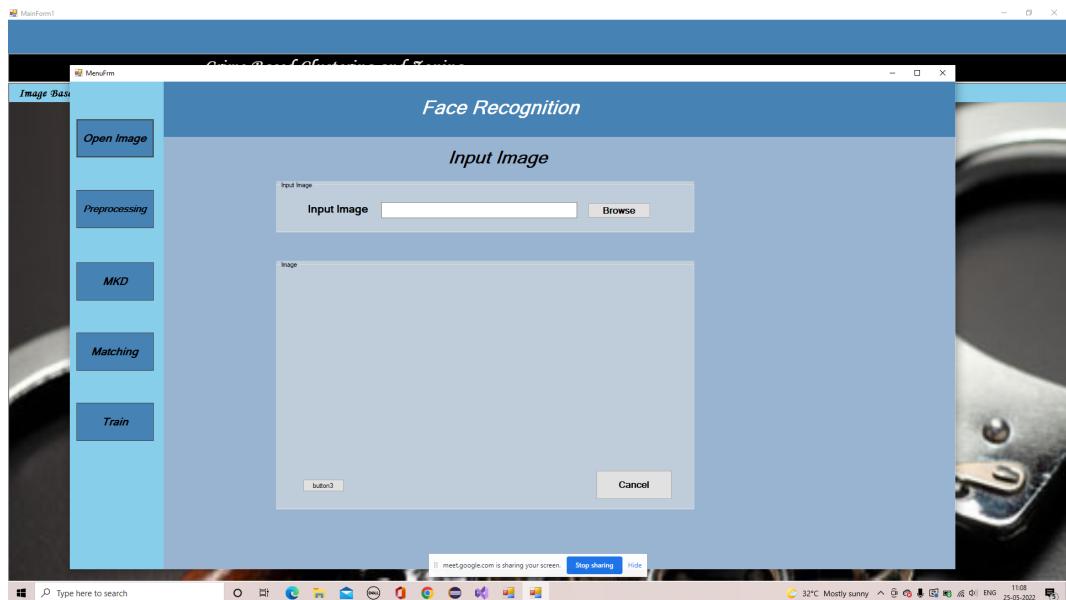


Figure 7.10: Image matching

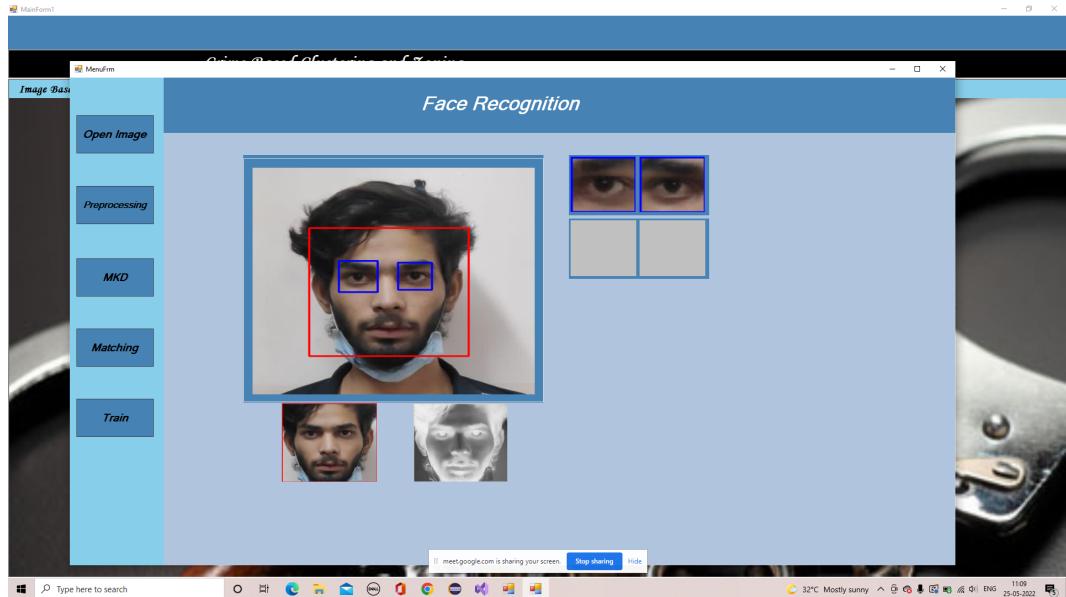


Figure 7.11: Feature matching

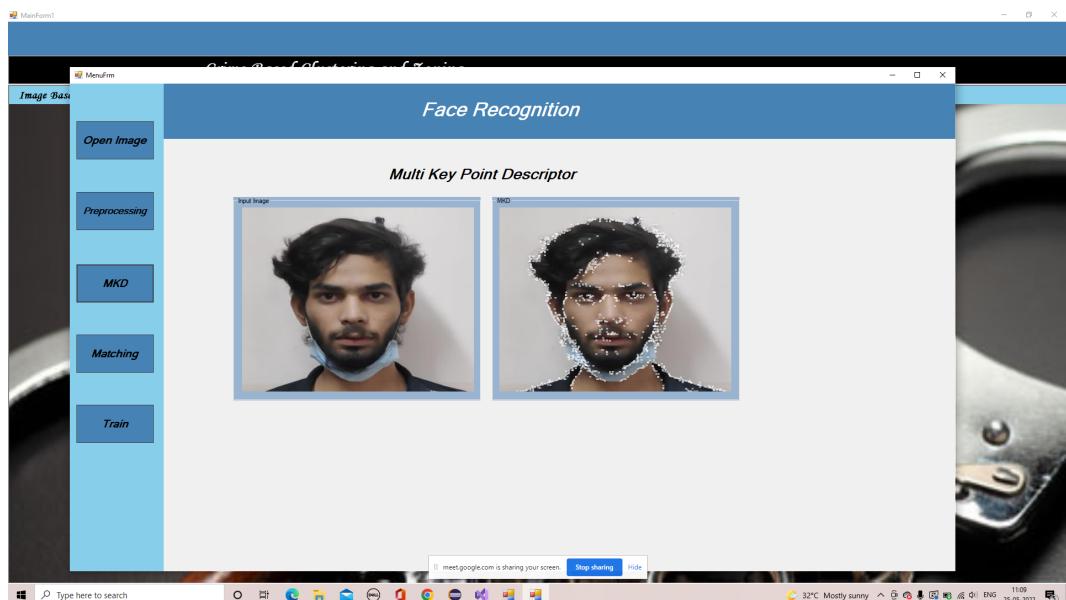


Figure 7.12: Key descriptor points

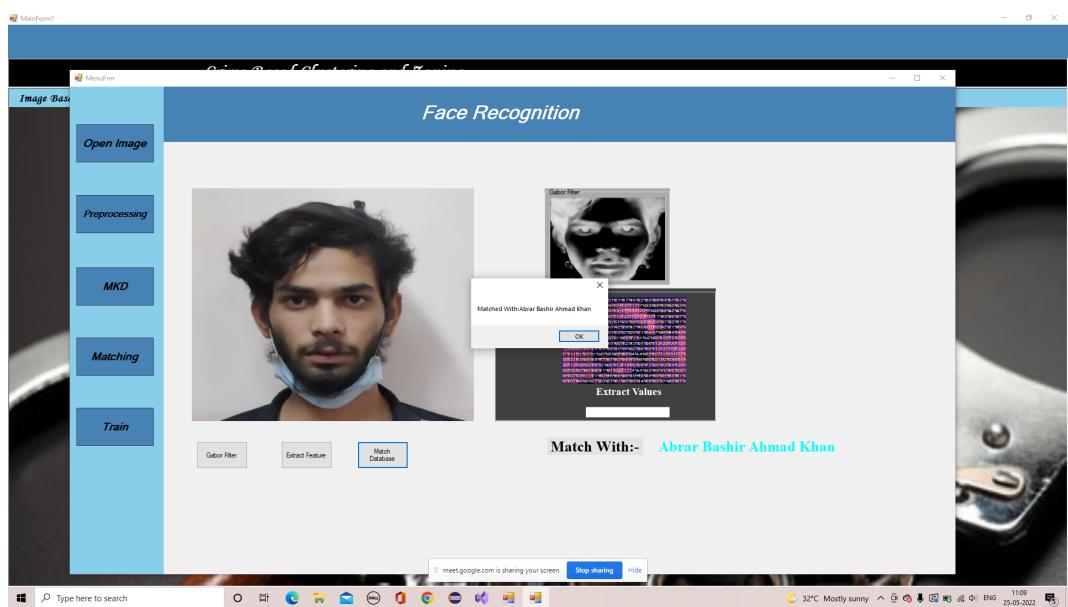


Figure 7.13: Image Matching results

CHAPTER 8

CONCLUSION AND FUTURE WORK

8.1 CONCLUSION

Crimes in India are rising at an alarming rate because of the factors such as 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 to identify 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 a easy to understand and comprehending way for better implementation of law enforcement and will also help in detection of a specific individual at the crime scene.

8.2 FUTURE WORK

This system is designed specially for the general public to be up-to-date about their surrounding's 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 king of law enforcement. This will help control the crime rates and avoid harm to the public.

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.
- [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
- [4] Mehmet Sevri, HacerKarakacan, 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
- [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*"

ANNEXURE A

PLAGIARISM REPORT

Document Information

Analyzed document	COMP_PID24_V1.docx (D137953901)
Submitted	2022-05-25T12:08:00.0000000
Submitted by	
Submitter email	mjpatil@kkwagh.edu.in
Similarity	19%
Analysis address	kpbirla.kkwagh@analysis.ouriginal.com

Sources included in the report

W	URL: http://cloudstechnologies.in/cloudtech-admin/basepaperfiles/1593149080crime%20rate%20prediction%20using%20k-nn.pdf	 1
Fetched:	2021-11-30T21:33:43.0430000	
W	URL: https://www.researchgate.net/publication/316443671_Cluster_based_zoning_of_crime_info	 1
Fetched:	2020-11-16T05:38:21.7570000	
W	URL: http://www.iosrjen.org/Papers/vol10_issue1/Series-3/J1001035659.pdf	 1
Fetched:	2021-09-09T12:45:14.7800000	
SA	ICPS_2018_paper_20.pdf Document ICPS_2018_paper_20.pdf (D36729338)	 1
W	URL: https://www.ijritcc.org/index.php/ijritcc/article/view/1392	 1
Fetched:	2022-05-25T12:11:33.3870000	

Entire Document

84%

MATCHING BLOCK 1/5

W

<http://cloudstechnologies.in/cloudtech-admin/b...>

Abstract Crime is one of the most predominant and alarming aspects in 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 reasons, factors and relations between 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.

95%

MATCHING BLOCK 2/5

W

<https://www.researchgate.net/publication/31644...>

We will 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.

98%

MATCHING BLOCK 3/5

W

http://www.iosrjen.org/Papers/vol10_issue1/Ser...

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

100%

MATCHING BLOCK 4/5

SA

[ICPS_2018_paper_20.pdf \(D36729338\)](ICPS_2018_paper_20.pdf (D36729338))

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].

PROBLEM STATEMENT

•

50%

MATCHING BLOCK 5/5

W

<https://www.ijritcc.org/index.php/ijritcc/arti...>

To study reasons, factors and relations between occurrence of different crimes and finding the most appropriate ways to control and avoid more crimes. •

To identify criminal pattern to pace up the mechanism of crime solutions. • To learn and identify a person using his facial images

Goal and Objectives:

- To classify clustered crimes based on frequency of crime occurrence during different years. • To find the difference between the facial characteristic of criminal and non- criminal faces. • To execute crime analysis from huge criminal data, an appropriate scientific field needs to be chosen. 2.1.2 Assumption and Scope • Data mining is used extensively in terms of analysis, investigation and discovery of patterns for occurrence of different crimes. • We propose a theoretical model based on data mining techniques such as clustering and classification to dataset

METHODOLOGY

- In the data pre-processing step, the attributes in the dataset are examined by 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 type and place of the crime occurrence • Custom map of India will be

created which displays these cluster zones in different colour coded sectors for better and easier understanding. • Image processing will be performed as image of a suspect will be the input in the system. System will analyse 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.

OUTCOME

- To understand the concept of data mining and machine learning which can be used for finding criminal patterns and behaviors.

TYPE OF PROJECT

- The project Crime Based Clustering and Zoning is application oriented. • The Crime Based Clustering and Zoning comes under the domain of Data Mining and Image processing.

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 a 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 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, bar graph is shown with 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 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 criminal if present in the input photo. It is then matched with the criminal dataset already fitted in the system.

CONCLUSION

Crimes in India are rising at an alarming rate because of the factors such as 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 to identify 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 a easy to understand and comprehending way for better implementation of law enforcement and will also help in detection of a specific individual at the crime scene.

FUTURE WORK

This system is designed specially for the general public to be up-to-date about their surrounding's 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.

Hit and source - focused comparison, Side by Side

Submitted text	As student entered the text in the submitted document.
Matching text	As the text appears in the source.

1/5	SUBMITTED TEXT	62 WORDS	84% MATCHING TEXT	62 WORDS
	<p>Abstract Crime is one of the most predominant and alarming aspects in 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 reasons, factors and relations between occurrence of different crimes and finding the most appropriate ways to control and avoid more crimes.</p>		<p>Abstract—Crime is one of the most predominant and alarming aspects in 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 becomes necessary to study various reasons, factors and relationship between different crimes that are occurring and finding the most appropriate methods to control and avoid more crimes.</p>	
<p>W http://cloudstechologies.in/cloudtech-admin/basepaperfiles/1593149080crime%20rate%20prediction%20.pdf</p>				

2/5	SUBMITTED TEXT	32 WORDS	95% MATCHING TEXT	32 WORDS
	<p>We will 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.</p>		<p>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.</p>	
<p>W https://www.researchgate.net/publication/316443671_Cluster_based_zoning_of_crime_info</p>				

3/5	SUBMITTED TEXT	27 WORDS	98% MATCHING TEXT	27 WORDS
	<p>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</p>		<p>to cluster the crime activities based on some predefined cases and the results of these clustering techniques are compared to find the best suitable clustering algorithm for</p>	
<p>W http://www.iosrjen.org/Papers/vol10_issue1/Series-3/J1001035659.pdf</p>				

4/5	SUBMITTED TEXT	22 WORDS	100% MATCHING TEXT	22 WORDS
	<p>to raise people's awareness regarding the dangerous locations and to help agencies to predict future crimes in a specific location within</p>			
<p>SA ICPS_2018_paper_20.pdf (D36729338)</p>				

5/5

SUBMITTED TEXT

23 WORDS

50% MATCHING TEXT

23 WORDS

To study reasons, factors and relations between occurrence of different crimes and finding the most appropriate ways to control and avoid more crimes. •

to study various reasons, factors and relationship between different crimes are occurring and finding the most appropriate methods to control and avoid more crimes.

 <https://www.ijritcc.org/index.php/ijritcc/article/view/1392>