GEO PROFILING MAPS BASED ON CRIME

Siddhant Magow, Harshit Kapoor, Chiranjeev Singh Bindra, Deepak Gupta
UG Student (Maharaja Agrasen Institute of Technology), UG Student (Maharaja Agrasen Institute of Technology), UG Student (Maharaja Agrasen Institute of Technology), Assistant Professor (Maharaja Agrasen Institute of Technology)
siddhant.magow@gmail.com, kapoor.harshit97@gmail.com, bindrachiranjeev@gmail.com, deepakgupta@mait.ac.in

Abstract— **Security has become an essential aspect of everyday life in India. Crimes are occurring rampantly even in broad daylight and its becoming difficult to predict the place and time of their occurrence due to their sudden nature. One of the ways to provide security is by enabling people with vital information on the security aspects for the route they would take during a travel from one place to another. This creates vigilance in traveler and precautionary measures could be taken to prevent any mishap. This project tries to envision an application that will suspect any possible security lapse in a location enroute a destination. A user would use a map which would display all possible areas of red alert for security from source to destination. Given a map (like Google Maps), the security-based profiling of various geographical routes traversed could be highlighted as red/yellow/green indicating high, medium and low risk zones respectively. This would enable the user to be vigilant while travelling through high risk zones in the map. The profiling of routes can be done on crime data obtained from police (Delhi Police in this case) website for arrested persons, their crime of conviction, place of crime committed, place of arrest etc. Machine learning algorithms could be used to learn the security features that contribute to the security of the location. Clustering algorithms can be used for this purpose. K means clustering is one of the simplest clustering algorithms which categorize the items into k groups of similarity. To calculate that similarity, it uses some distance metric (Euclidean, Manhattan, etc.) for distance measurement.**

Index Terms—**Crime Prediction, K-Means, Clustering, Data Mining, Crime Prone Areas.**

I. INTRODUCTION

Criminals are nuisance for the general public in the entire world for quite a while now and measures are required to reduce the crimes that are being committed. Current policing procedures aim at finding the lawbreakers, fundamentally after the crime has been committed. However, with the assistance of technology, we can utilize historic crime data to identify crime patterns and use them to predict crimes before they are committed. The proposed system uses machine learning algorithms to predict crime prone areas. The focus of the project is on crime occurring in the NCT of Delhi and hence the dataset has been obtained from the Delhi Police website but the procedure followed in this project can be easily extended to crime dataset of any region. The dataset consists of details of the crimes committed on a daily basis which is uploaded on the Delhi Police website daily. The dataset also consists of the region where crime was committed, the address of the crime, the act under which the crime falls, the time of occurrence of the crime, FIR report number etc. Out of these only a few of these features are useful to the project. However, the data was available in non-searchable pdf format along with some missing and erroneous values and hence had to be pre-processed to be able to use it in the project. After pre-processing the resulting dataset was available in Comma Separated Values (CSV) format which was suitable for applying the machine learning algorithms. Clustering algorithms have been used in the project. Clustering techniques are used to form clusters from the dataset which are then analyzed for determining crime prone areas. Each cluster represents group of crimes which are nearest to it. These clusters are represented on the map of Delhi. Clusters store location where crimes have been committed. These clusters are classified according to the frequency of crime data points in the cluster. High populated clusters become high risk zones whereas clusters with sparsely populated clusters become medium risk and low risk zones on the basis of their cluster density. Preventive measures are implemented on the basis of the most prevalent crime in each area. K-means is the most widely used clustering algorithm in machine learning due to its simplicity. It is suitable for clustering large data sets as it has smaller time and space complexity. It therefore finds successful application in various areas, including computer vision, geostatistics, astronomy and agriculture. The system uses clustering over any other supervised technique such as classification, since crimes that are committed vary in nature and there can also be unsolved crimes in the crime database. Therefore, classification techniques which need labelled data of the solved crimes for learning, will not give good results while predicting future crimes. To summarize the system tries to achieve the following objectives:

➢ To profile various areas of the city based on crime and identify the potential high-risk zones.

➢ To help the users to take necessary precautions who are planning a route through these high-risk zones.

➢ To help the security forces in security profiling of areas and help them to take necessary and swift actions in areas prone to crime.

➢ To help the users identify the nature of crime and take appropriate steps while travelling on that route.

➢ To apply Machine learning concepts to smoothen and automate the law enforcement process.

The remainder of the paper is divided as follows: Section II reviews the related work performed in this field. Section III and IV give the methodology and implementation

Received: 15 May 2019/Accepted 05 July 2019.

respectively. Section V summaries the results which were obtained and Section VI give the conclusion and future scope of this paper.

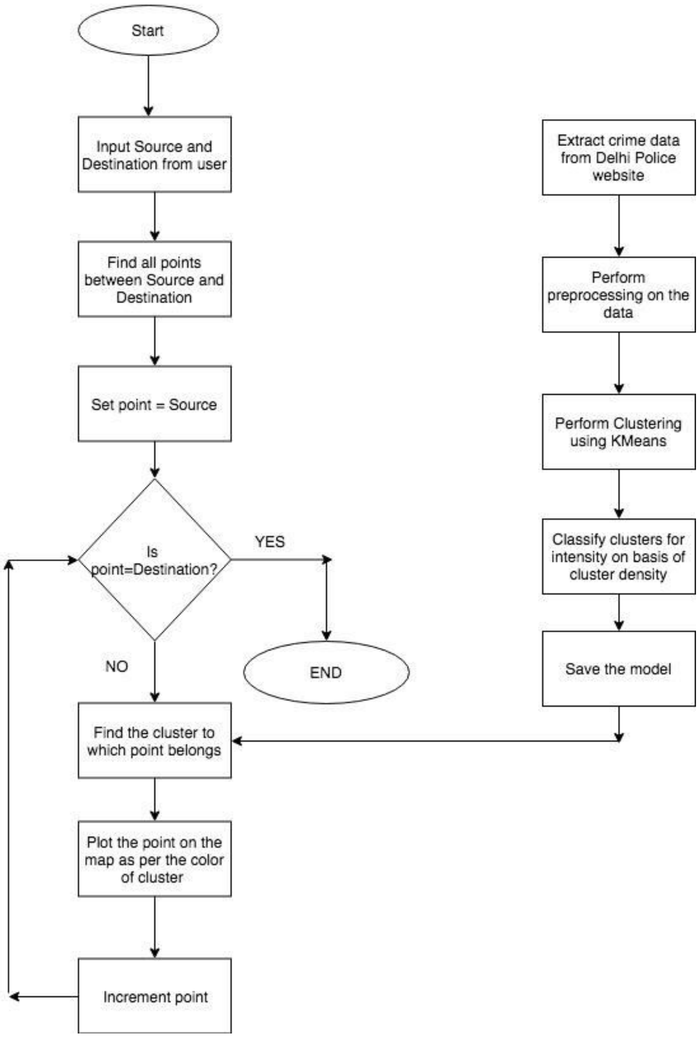
II. RELATED WORK

Criminology is the field of studying crime characteristics and ensuring law enforcement. It is one of the most significant applications of data mining techniques. Research is going on simulation models on numerous theoretical and empirical models of crime [1]. In the recent past, there has been a tremendous increase in crime. As a result, Controlling and Monitoring crime has become a difficult task. There is a need to use the cutting-edge technology to create a user interactive system that reduce the burden of our police force [2]. Various researches have been conducted to help data scientists to discover new patterns and relationships in the occurrence of crimes so that they can be predicted before they are committed. The classification is based mainly on various crime characteristics such as crime type, location, time etc. The results of both classifications and Clustering can be used for prediction of trends and behavior of the given objects (Crimes and Criminals) [3]. Each cluster represents group of crimes which are nearest to it. These clusters are represented on the map of Delhi. Such clusters are useful in identifying a crime pattern or a crime spree [4]. Using this, the police force can understand the crime pattern across a certain region, or interval of time is, and thus can take proactive action to prevent those crimes from occurring. This would save them a lot of time, money and effort. The current systems extract the data and run appropriate algorithms on data. The output is presented to the user in the form of clusters using a clustering algorithm like K-means clustering algorithm [5]. De Bruin et. al. [6] introduced a framework for crime trends using a new distance measure for comparing all individuals based on their profiles and then clustering them accordingly. In Malaysia, Nazlena Mohamad Ali et al.[7] wrote about the development of a Visual Interactive Malaysia Crime News Retrieval System (i-JEN) .Their main objectives were to model crime-based events and using them to train the clustering model. The model will then be integrated into a usable and robust system that will contribute to the better understanding of the crime data consumption in the Malaysian context. Jyoti Aggarwal et al. [8] have proposed a system to use k means algorithm to extract essential information from the large crime datasets and using them to develop a model which will assists the police to identify and analyse crime patterns to reduce further occurrences of similar crimes and provide information to help develop proactive strategies for combating crime. Sutapat Thiprungsri [9] examined the application of cluster analysis in the accounting domain, particularly discrepancy detection in audit. Their focus was on examining the use of clustering technology to automate fraud International Journal of Computer Applications (0975 – 8887) Volume 83 – No4, December 2013 2 filtering during an audit. They used cluster analysis to help auditors focus their efforts when evaluating group life insurance claims. Tony H. Grubesic [10] explored the use of a generalized partitioning method known as fuzzy clustering for hot-spot detection. He analysed behavioral and visual difference between fuzzy clustering and two hard clustering techniques (medoid and k-means), across a range of cluster values. His empirical results suggested that a fuzzy clustering approach gives better results and can handle the outliers efficiently. D.K. Tayal et al.[11] proposed an method for developing a crime detection and criminal identification system for Indian cities using data science. Their approach consists of six modules, namely—data extraction (DE), data pre-processing (DP), clustering, Google map representation, classification and WEKA® implementation. The first module gathers the unstructured crime data from various sources across the web. The second module, DP performs data preprocessing i.e. it cleans, integrates and reduces the crime data into structured format with 5,038 crime instances which were represented using 35 predefined crime attributes. Rest four modules were used for crime detection, criminal identification and prediction, and crime verification, respectively. T.Wang et al.[12] propose a pattern detection algorithm called Series Finder, that derives useful crime patterns from the crime data in the database, starting from a “seed” of a few crimes. Series Finder takes into account both the common as well as the unique aspects of each pattern, and shows promising results on a decade’s worth of crime pattern data collected by the Crime Analysis Unit of the Cambridge Police Department. Qusay Bsol et al.[13] proposed two major sequential stages in Document Clustering “Extraction Features and Clustering Algorithms” as well as the major challenges and the key issues in designing extraction features and clustering algorithms. Their approach assisted the law enforcement officers and detectives to enhance performance and speed up the process of solving crimes. Bashar Aubaidan et. Al[14] presented the results of an experimental study of two document clustering techniques namely kmeans and k-means++. They compared the two main approaches in crime document clustering. They proposed that the k-means++ clustering algorithm can be used to identify best seed for initial cluster centres that are derived from the crime data. They conducted a comparative study of two main clustering algorithms, namely k-means and k-means++. The method of their study included a pre-processing phase, which in turn involved tokenization, stop-words removal and stemming. In addition, they evaluated the impact of the two similarity/distance measures (Cosine similarity and Jaccard coefficient) on the results of the two clustering algorithms. Experimental results on multiple crime data sets showed that k-mean++ can significantly (with the significance interval at 95%) work better than k-means by finding the best seed for initial cluster centres. Mohammed Reza Keyvanpour et al.[15] used an approach based on data mining techniques to extract important information from police narrative reports which are written in plain text. With the help of this approach, crime data was automatically stored into a database. They also applied a SOM clustering method in the scope of crime analysis and used the clustering results in order to perform crime matching process. Keeping in mind all the previous research and projects which have been done and implemented, the paper proposes a system which provides information to the user related to the crime based on the route he/she decides to take while travelling from one location to another in the city. The system will refer to the results of the previous research work done to profile various crime hotspots in a city and then apply it specifically for the route a user takes in the city. It will also make the dataset specific to a city i.e. Delhi (in this case) and also profile the areas for the same. The system will use ML algorithms like Clustering to perform these operations and use a system to update the dataset on daily basis as it’s done on the police website.

Received: 15 May 2019/Accepted 05 July 2019.

III. Methodologies

Below is a flow chart explaining the working of the system:


Start
Input Source and Destination from user
Find all points between Source and Destination
Set point = Source
Is point=Destination?
YES
NO
END
Find the cluster to which point belongs
Plot the point on the map as per the color of cluster
Increment point
Extract crime data from Delhi Police website
Perform preprocessing on the data
Perform Clustering using KMeans
Classify clusters for intensity on basis of cluster density
Save the model

This paper proposes an automated system to identify the potentially dangerous regions according to the level of criminal activity along a route between a source and destination which is taken from a user as an input. It uses the concepts of machine learning and clustering to form various clusters on the map (18 to be exact). These clusters are formed on the map according to the location of occurrence of crimes and are of varying density i.e. some have high density of criminal activity and others have moderate to low criminal activity. The data used in this project is obtained from crime data available on the Delhi Police Website. This data was used to extract the place of occurrence of crime and convert it to latitude and longitude through geocoding api and these coordinates were then plotted on a Map. Once clustering is complete, we obtain the cluster centres. Using the clusters centres, we divided the clusters into three categories: 1) High Risk 2) Medium Risk 3) Low Risk. This classification was made on the basis of density of crimes in each cluster. Then using the directions api, a route was obtained between the source and destination which was shown on the map. Then various points along this route were obtained and each point on the route was checked to find whether it lies in a high risk, medium risk or a low risk cluster and accordingly it's plotted on the map in red, yellow or green colour respectively. Thus we get a route composed of regions coloured in red, yellow or green which is shown to the user along with an alert message which tells the user how much percent of the route falls in high risk, moderate risk and least risk category. Thus, the user can plan their trip accordingly. Below are the processes which have to be carried out before implementing the front end and backend:

1) DATA COLLECTION: The data for the project was obtained from the Delhi Police Website. This data however was in nonsearchable pdf format. This format was unsuitable for use in Machine Learning. Thus, it had to be converted into searchable pdf format. For this purpose, each pdf was converted into a series of .jpeg images and then these images were converted into .csv files using OCR editor.

2) DATA PRE-PROCESSING: In pre-processing, each .csv file obtained in the above step was read into a pandas dataframe and the unwanted newline characters were removed. Then all these dataframes were merged into a single dataframe to obtain the dataset for the project.

3) GEOCODING: The address attribute in the dataset was replaced by latitude and longitude using Bing GeoCoder API. After obtaining the Latitude and Longitude, the address attribute was dropped.

4) CRIME CLASSIFICATION: The Section attribute in the dataset was replaced by its corresponding crime by using the section number and the act involved. After obtaining the crime, the section attribute was dropped.

5) DATA PARTITIONING: The dataset was partitioned into day crime dataset and night crime dataset using the time at which the crime had occurred. After partitioning the dataset, the timestamp attribute was dropped from both the partitioned datasets.

6) CLUSTERING: The outliers (points which lied outside Delhi) were removed from the partitioned datasets and clustering was performed using K means algorithm with k=18 as there are a total of 18 police jurisdictional districts in Delhi. The training data of K means algorithm comprised only of Latitude and Longitude attributes.

7) SAVING THE MODEL: After performing clustering on the day and night crime datasets, their respective K means models were saved using Pickle module of Python. These saved models were later used in the main backend file of the project.

After carrying the above processes, the front and backend of the project were implemented. Given below is a description of the front end and backend used in the project:

FRONT END: The front end consists of an html file with a map in the background obtained from Google Maps JavaScript API and two text boxes which take the source and destination from the user respectively. It also consists of a button which is used to fetch the source and destination and display a route to the user which has been profiled on the basis of crime. The following processes take place when the user clicks on the button after entering the source and the destination:

Received: 15 May 2019/Accepted 05 July 2019.

➢ All the points between the source and destination are found out using Google Directions API and stored in an array called pointsArray.

➢ The current time is calculated and an AJAX request is made to the python backend with the pointsArray and the current time which are sent in JSON format.

➢ In the success callback of the AJAX request,

o The response from the backend is converted into JSON format (an array of JSON objects as the response for multiple points on the route is being received).

o For each item in the JSON response Array, the colour is found out and the corresponding colour's count is incremented. Also, the crime associated with the point is stored in the corresponding colour's crime list.

o After iterating for each item in the response array, the percentage of the route covered by each colour i.e. the percentage of route covered under different intensity categories of crime (Red: Very Dangerous, Yellow: Moderate, Green: Safe) is calculated. Also, the most prevalent crime for each intensity category is found out.

o An alert is displayed to the user displaying the above information and a map with the route geo profiled according to the crime is opened in a new tab.

BACK END: After receiving the request from the front-end following processes take place:

➢ The model corresponding to the time received from the AJAX request is loaded.

➢ Data points received in the request are extracted and stored in a list.

➢ The model is used to perform clustering on the corresponding crime data and the result is stored in a variable called prediction.

➢ For each cluster centre, the most prevalent crime is found and stored in a list

➢ The no. of points belonging to each cluster are found and stored in a list.

➢ Using the above list, the range from minimum to maximum value is divided into 3 intervals and all the clusters are assigned a colour according to the interval to which they belong.

➢ For each data point in the list,

o Clustering is performed on the point using loaded model.

o The point is assigned a colour according to the colour of its corresponding cluster found out in the previous step. Also, the most prevalent crime of the cluster is assigned to the point.

o The colour and crime of the point is stored in a dictionary with keys 'colour' and 'crime'.

o This dictionary is appended to the final result list.

➢ All the points are plotted on a map according their corresponding colours using gmplot module of the python.

➢ The final result list is converted to JSON format and the response is returned to the front end.

INTERFACE BETWEEN FRONT END AND BACKEND: The interaction between front end and backend is done using Flask Module of Python. Flask is a python micro framework that helps in interaction by creating a local host server. On the root, the front-end html file is rendered and the request made by the user for finding the route are handled by an onClick handler which sends and AJAX request to the local host server at the end point receiver. The data received is manipulated and the response is sent back to the front creating a bidirectional connection between backend and front end.

IV. Implementation

K Means Algorithm: Below is an algorithm for K Means clustering which has been used in this project:

1) Let X = {x1,x2,x3,……..,xn} represent the set of data points and let V = {v1,v2,…….,vc} represent the set of centres.

2) Randomly select 'c' cluster centres from the set of data points.

3) Calculate the distance between each data point in X and cluster centres V chosen in the previous step.

4) Assign the data point to the cluster whose distance from the point is minimum amongst all the cluster centres.

$$v_i = (1/c_i) \sum_{j=1}^{c_i} x_i$$

5) Recalculate the new cluster centre using the above formula: Where, 'ci' represents the number of data points in ith cluster.

6) Repeat the algorithm from step 3) till the clusters become stable and no data point is reassigned.

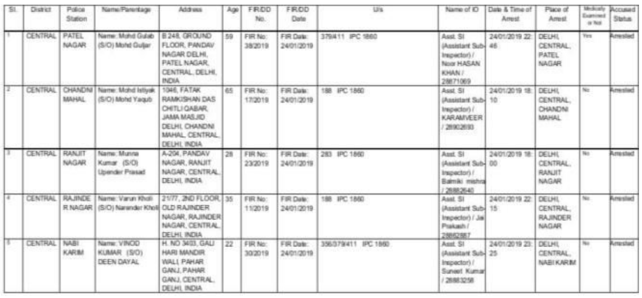
Experimental Setup: The project proposes a web application written in JavaScript which can be run easily on any Java Script Enabled browser. The backend of the project is written in Python which is the go to language these days for implementing machine learning algorithms. Also, external APIs like Google Directions API, Maps JavaScript API and Bing Maps Geocoding API have been used for various applications like rendering the route between source and destination, Displaying Maps Background and plotting points on the map respectively. On the front end no special hardware

Received: 15 May 2019/Accepted 05 July 2019.

or software requirements are needed to run the application. On the server side, all the necessary machine learning modules should have been installed and an appropriate processor may be needed if the dataset is large. Also it is assumed that the application would be used in the NCT of Delhi as the dataset of Delhi is used.

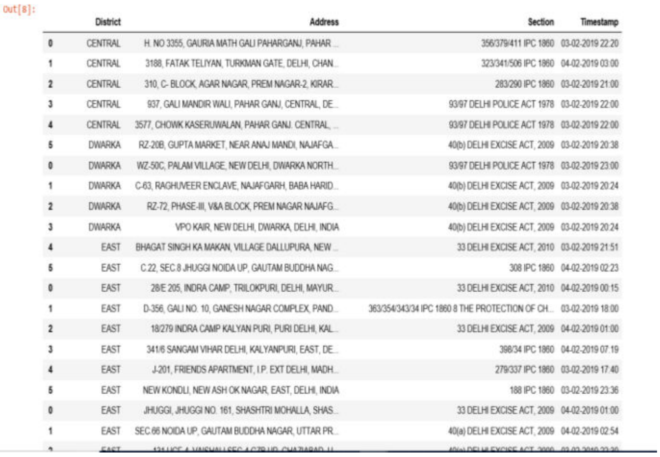
Input Parameters: The input parameters are the source and destination which are taken as an input from the user. The project provides a route between the source and destination along with crime information about that route.

Data Pre-Processing: Since the data was available in a non-searchable pdf, the project made use of Abby Fine Reader, a popular ocr text extractor, for extracting the relevant data from the pdf. The project first divided the original pdf into a number of images (.jpg format) using pdf2html module of python. Then the ocr text editor of Abby Fine Reader was used to extract data from these images in csv format. The csv files obtained from different images were pre-processed and then merged into a single csv file. In pre-processing, the unwanted attributes from the csv files were removed and then unwanted new line characters were replaced with spaces. From the Address attribute we created two new attributes, Latitude and Longitude, using Bing Maps Geo Coding API and dropped the Address attribute. After removing the null values arising from the above steps, our dataset was finally ready. Below is the raw data obtained from Delhi Police Website.



Sl.	District	Police Station	Name/Parentage	Address	Age	FIR/DD No.	FIR/DD Date	U/s	Name of IO	Date & Time of Arrest	Place of Arrest	Medically Examined or Not	Accused Status
1	CENTRAL	PATEL NAGAR	Name: Mohd Gulab (S/O) Mohd Guljar	B 248, GROUND FLOOR, PANDAV NAGAR DELHI, PATEL NAGAR, CENTRAL, DELHI, INDIA	59	FIR No: 38/2019	FIR Date: 24/01/2019	379/411 IPC 1860	Asst. SI (Assistant Sub-Inspector) / Noor HASAN KHAN / 28871069	24/01/2019 22:46	DELHI, CENTRAL, PATEL NAGAR	Yes	Arrested
2	CENTRAL	CHANDNI MAHAL	Name: Mohd Istiyak (S/O) Mohd Yaqub	1046, FATAK RAMKISHAN DAS CHITLI QABAR, JAMA MASJID DELHI, CHANDNI MAHAL, CENTRAL, DELHI, INDIA	65	FIR No: 17/2019	FIR Date: 24/01/2019	188 IPC 1860	Asst. SI (Assistant Sub-Inspector) / KARAMVEER / 28902693	24/01/2019 18:10	DELHI, CENTRAL, CHANDNI MAHAL	No	Arrested
3	CENTRAL	RANJIT NAGAR	Name: Munna Kumar (S/O) Upender Prasad	A-204, PANDAV NAGAR, RANJIT NAGAR, CENTRAL, DELHI, INDIA	28	FIR No: 23/2019	FIR Date: 24/01/2019	283 IPC 1860	Asst. SI (Assistant Sub-Inspector) / Balmiki mishra / 28882640	24/01/2019 18:00	DELHI, CENTRAL, RANJIT NAGAR	No	Arrested
4	CENTRAL	RAJINDER NAGAR	Name: Varun Kholi (S/O) Narender Kholi	21/77, 2ND FLOOR, OLD RAJINDER NAGAR, RAJINDER NAGAR, CENTRAL, DELHI, INDIA	35	FIR No: 11/2019	FIR Date: 24/01/2019	188 IPC 1860	Asst. SI (Assistant Sub-Inspector) / Jai Prakash / 28862887	24/01/2019 22:15	DELHI, CENTRAL, RAJINDER NAGAR	No	Arrested
5	CENTRAL	NABI KARIM	Name: VINOD KUMAR (S/O) DEEN DAYAL	H. NO 3403, GALI HARI MANDIR WALI, PAHAR GANJ, PAHAR GANJ, CENTRAL, DELHI, INDIA	22	FIR No: 30/2019	FIR Date: 24/01/2019	356/379/411 IPC 1860	Asst. SI (Assistant Sub-Inspector) / Suneet Kumar / 28883258	24/01/2019 23:25	DELHI, CENTRAL, NABI KARIM	No	Arrested

Fig1. Data in non searchable format as obtained from Delhi police website

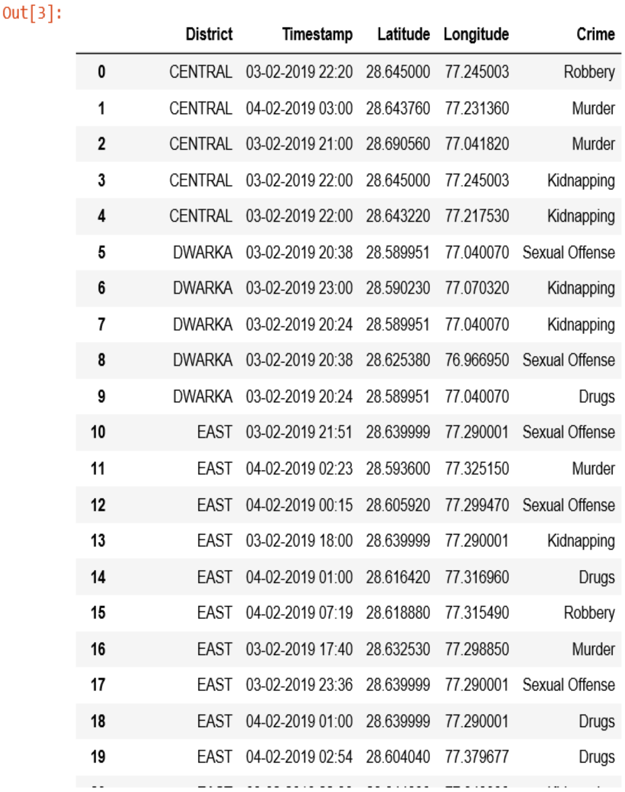


Out[8]:

	District	Address	Section	Timestamp
0	CENTRAL	H. NO 3355, GAURIA MATH GALI PAHARGANJ, PAHAR ...	356/379/411 IPC 1860	03-02-2019 22:20
1	CENTRAL	3188, FATAK TELIYAN, TURKMAN GATE, DELHI, CHAN...	323/341/506 IPC 1860	04-02-2019 03:00
2	CENTRAL	310, C- BLOCK, AGAR NAGAR, PREM NAGAR-2, KIRAR...	283/290 IPC 1860	03-02-2019 21:00
3	CENTRAL	937, GALI MANDIR WALI, PAHAR GANJ, CENTRAL, DE...	93/97 DELHI POLICE ACT 1978	03-02-2019 22:00
4	CENTRAL	3577, CHOWK KASERUWALAN, PAHAR GANJ. CENTRAL, ...	93/97 DELHI POLICE ACT 1978	03-02-2019 22:00
5	DWARKA	RZ-20B, GUPTA MARKET, NEAR ANAJ MANDI, NAJAFGA...	40(b) DELHI EXCISE ACT, 2009	03-02-2019 20:38
0	DWARKA	WZ-50C, PALAM VILLAGE, NEW DELHI, DWARKA NORTH...	93/97 DELHI POLICE ACT 1978	03-02-2019 23:00
1	DWARKA	C-63, RAGHUVEER ENCLAVE, NAJAFGARH, BABA HARID...	40(b) DELHI EXCISE ACT, 2009	03-02-2019 20:24
2	DWARKA	RZ-72, PHASE-III, V&A BLOCK, PREM NAGAR NAJAFG...	40(b) DELHI EXCISE ACT, 2009	03-02-2019 20:38
3	DWARKA	VPO KAIR, NEW DELHI, DWARKA, DELHI, INDIA	40(b) DELHI EXCISE ACT, 2009	03-02-2019 20:24
4	EAST	BHAGAT SINGH KA MAKAN, VILLAGE DALLUPURA, NEW ...	33 DELHI EXCISE ACT, 2010	03-02-2019 21:51
5	EAST	C.22, SEC.8 JHUGGI NOIDA UP, GAUTAM BUDDHA NAG...	308 IPC 1860	04-02-2019 02:23
0	EAST	28/E 205, INDRA CAMP, TRILOKPURI, DELHI, MAYUR...	33 DELHI EXCISE ACT, 2010	04-02-2019 00:15
1	EAST	D-356, GALI NO. 10, GANESH NAGAR COMPLEX, PAND...	363/354/343/34 IPC 1860 8 THE PROTECTION OF CH...	03-02-2019 18:00
2	EAST	18/279 INDRA CAMP KALYAN PURI, PURI DELHI, KAL...	33 DELHI EXCISE ACT, 2009	04-02-2019 01:00
3	EAST	341/6 SANGAM VIHAR DELHI, KALYANPURI, EAST, DE...	398/34 IPC 1860	04-02-2019 07:19
4	EAST	J-201, FRIENDS APARTMENT, I.P. EXT DELHI, MADH...	279/337 IPC 1860	03-02-2019 17:40
5	EAST	NEW KONDLI, NEW ASH OK NAGAR, EAST, DELHI, INDIA	188 IPC 1860	03-02-2019 23:36
0	EAST	JHUGGI, JHUGGI NO. 161, SHASHTRI MOHALLA, SHAS...	33 DELHI EXCISE ACT, 2009	04-02-2019 01:00
1	EAST	SEC.66 NOIDA UP, GAUTAM BUDDHA NAGAR, UTTAR PR...	40(a) DELHI EXCISE ACT, 2009	04-02-2019 02:54

Fig2. The dataset with minimum processing

Shown above is the dataset with minimum pre-processing i.e. unnecessary columns and new line characters dropped from the dataset.



Out[3]:

	District	Timestamp	Latitude	Longitude	Crime
0	CENTRAL	03-02-2019 22:20	28.645000	77.245003	Robbery
1	CENTRAL	04-02-2019 03:00	28.643760	77.231360	Murder
2	CENTRAL	03-02-2019 21:00	28.690560	77.041820	Murder
3	CENTRAL	03-02-2019 22:00	28.645000	77.245003	Kidnapping
4	CENTRAL	03-02-2019 22:00	28.643220	77.217530	Kidnapping
5	DWARKA	03-02-2019 20:38	28.589951	77.040070	Sexual Offense
6	DWARKA	03-02-2019 23:00	28.590230	77.070320	Kidnapping
7	DWARKA	03-02-2019 20:24	28.589951	77.040070	Kidnapping
8	DWARKA	03-02-2019 20:38	28.625380	76.966950	Sexual Offense
9	DWARKA	03-02-2019 20:24	28.589951	77.040070	Drugs
10	EAST	03-02-2019 21:51	28.639999	77.290001	Sexual Offense
11	EAST	04-02-2019 02:23	28.593600	77.325150	Murder
12	EAST	04-02-2019 00:15	28.605920	77.299470	Sexual Offense
13	EAST	03-02-2019 18:00	28.639999	77.290001	Kidnapping
14	EAST	04-02-2019 01:00	28.616420	77.316960	Drugs
15	EAST	04-02-2019 07:19	28.618880	77.315490	Robbery
16	EAST	03-02-2019 17:40	28.632530	77.298850	Murder
17	EAST	03-02-2019 23:36	28.639999	77.290001	Sexual Offense
18	EAST	04-02-2019 01:00	28.639999	77.290001	Drugs
19	EAST	04-02-2019 02:54	28.604040	77.379677	Drugs

Fig3. The final processed dataset.

Above is the final processed dataset which was fed to the model. Here the address has been converted to latitude and longitude and the sections have been replaced by their corresponding crimes.

V. RESULTS AND DISCUSSION

After applying machine learning algorithms on the processed data, we obtained the following results:

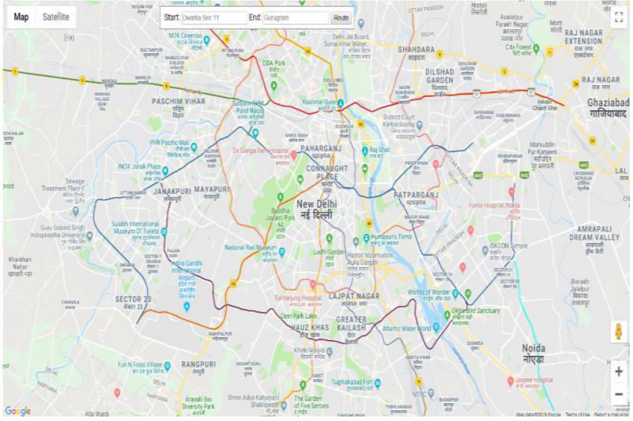


Fig4. Front end of the project

Shown above is how the front end of the project looks like. It consists of a map in the background along with two text boxes for taking source and destination as an input and a button which gives a geo profiled map with route between the source and destination.

Received: 15 May 2019/Accepted 05 July 2019.

The below figure shows the cluster centres of all the 18 clusters shown on the map with different colours.

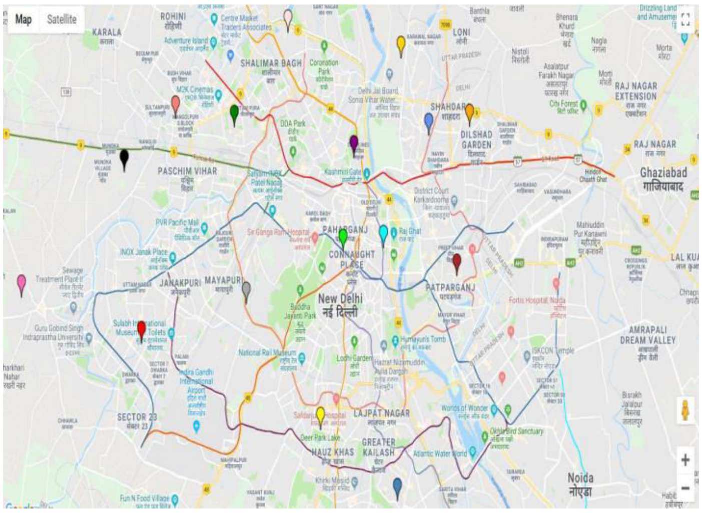
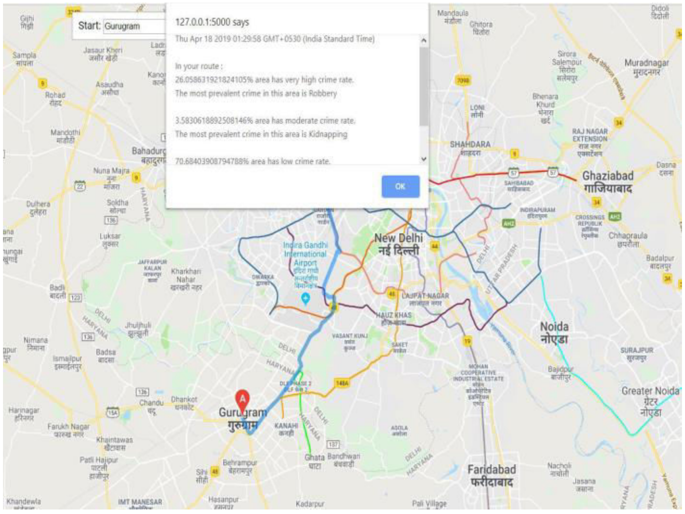


Fig5: Cluster centres of all 18 clusters.



Start: Gurugram

127.0.0.1:5000 says

Thu Apr 18 2019 01:29:58 GMT+0530 (India Standard Time)

In your route :

26.058631921824105% area has very high crime rate.

The most prevalent crime in this area is Robbery

3.5830618892508146% area has moderate crime rate.

The most prevalent crime in this area is Kidnapping

70.68403908794788% area has low crime rate.

Fig6. All 18 clusters coloured according to their crime density

The above Fig 6 shows all our cluster centres on the map coloured according to their density of crime, i.e. red for high intensity, yellow for moderate intensity and green for low intensity regions.

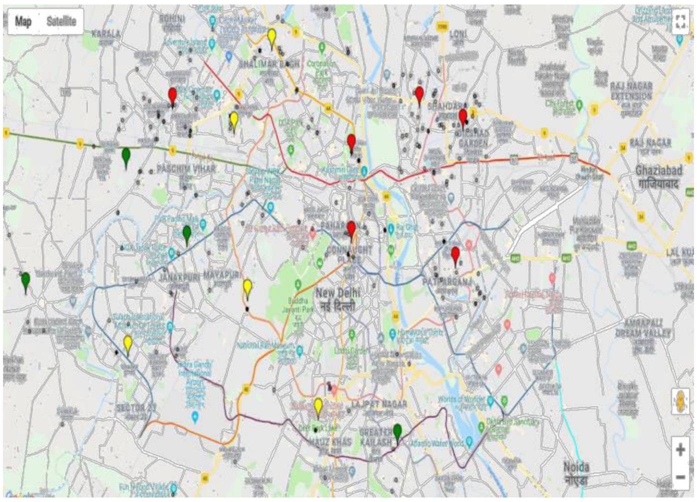


Fig7. Route between the source (A) and destination (B) along with an alert message about the crime along the route.

The above figure shows the route between the entered source and destination. After displaying the route an alert message is shown which tells the user the percentage of the route under different intensity zones and the most prevalent crime in those zones. Eg: 26% in the red zone (high crime rate) with robbery as the most prevalent crime.

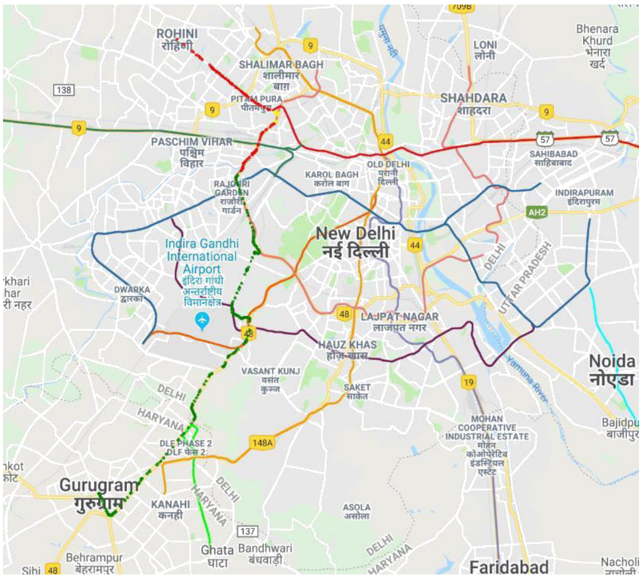


Fig8. Route between source and destination with points on the route coloured according the intensity of cluster they belong to.

Received: 15 May 2019/Accepted 05 July 2019.

Shown above in Fig 8, is the route between the Gurugram and Rohini. The route has points of different colours i.e red, yellow and green. These points are coloured according to the intensity of the cluster these points belong to, i.e red for high crime intensity, yellow for moderate and green for low crime intensity.

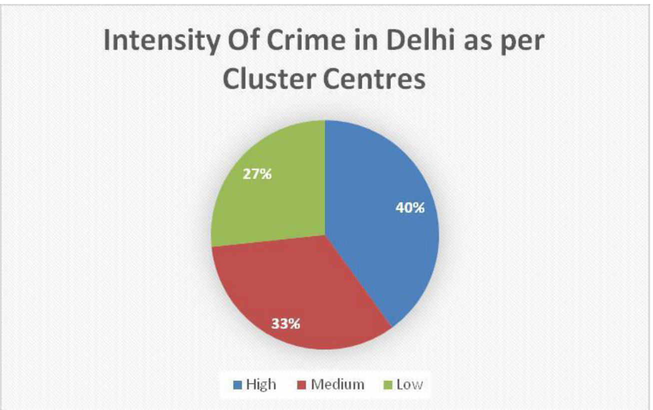


Fig9: Intensity of crimes In Delhi as per Cluster centres

In above Fig9, intensity of crimes as per cluster centres is shown. We see that 40% of clusters have high crime rates, 33% medium crime and 27% low crime rates.

The below figures show the geoprofiling crime intensity in a route from Kashmere Gate to Rohini, New Delhi.

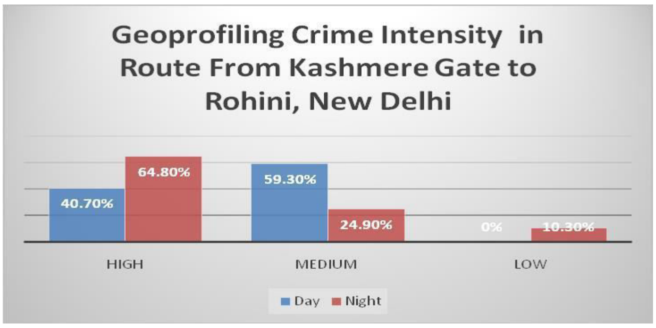
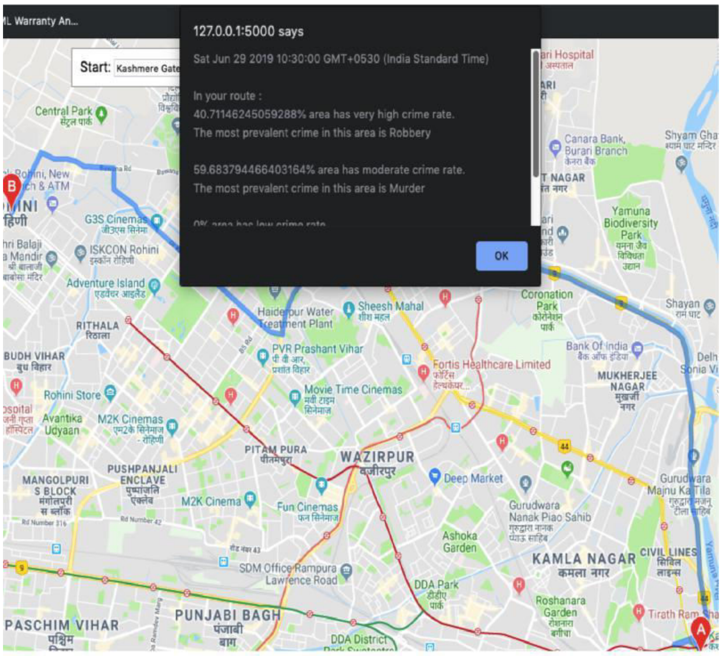


Fig10: Geo Profiling crime intensity in a route
Fig11: Route

.

Fig10 shows a comparison of crime rates along the shown route in day and night time while Fig11 and Fig12 show the route on map in day and night time respectively.

So to summarise our application provides a geo profiled route to the user with the information about the different intensity and most prevalent types of crime along that route. This will enable the users to plan their trip accordingly, know which areas have a high crime rate and plan what precautions must be taken before going on that route.



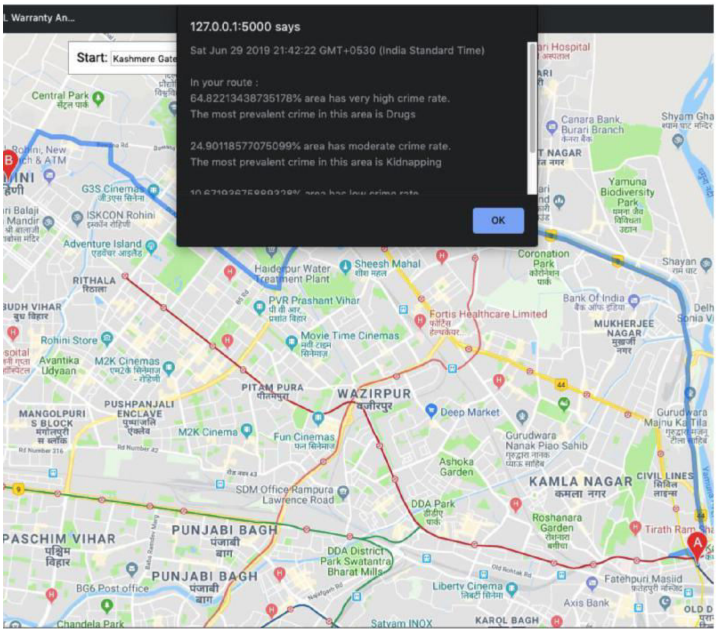
127.0.0.1:5000 says

Sat Jun 29 2019 10:30:00 GMT+0530 (India Standard Time)

In your route :
40.71146245059288% area has very high crime rate.
The most prevalent crime in this area is Robbery

59.683794466403164% area has moderate crime rate.
The most prevalent crime in this area is Murder

0% area has low crime rate



127.0.0.1:5000 says

Sat Jun 29 2019 21:42:22 GMT+0530 (India Standard Time)

In your route :
64.82213438735178% area has very high crime rate.
The most prevalent crime in this area is Drugs

24.90118577075099% area has moderate crime rate.
The most prevalent crime in this area is Kidnapping

10.67193675889328% area has low crime rate

VI. CONCLUSION AND FUTURE SCOPE

This project has proposed an application that suspects any possible security lapse in a location enroute a destination. A user can use a map which displays all possible areas of red alert for security from source to destination. Given a map (like Google Maps), the security based profiling of various geographical routes traversed is highlighted as red/yellow/green indicating high, medium and low risk zones respectively. This enables the user to be vigilant while travelling through high risk zones in the map. The profiling of routes has been done on crime data obtained from police (Delhi Police in this case) website for arrested persons, their crime of conviction, place of crime committed, place of arrest etc.

In the future, the application can be extended to show the best route (with least crime) out of all the possible routes between source and destination that the user enters. The data

Received: 15 May 2019/Accepted 05 July 2019.

collection and pre-processing step could also be automated so that the data for each data is extracted automatically from the website and merged with the final dataset after automatic pre-processing. The same application could be extended to other platforms like mobile devices as well. Also, the domain of the project could be increased covering more regions and states so that people belonging to different regions can use this application as well.

REFERENCES

[1] K. Zakir Hussain, M. Durairaj and G. Rabia Jahani Farzana, “Application of Data Mining Techniques for Analyzing Violent Criminal Behaviour by Simulation Model”, International Journal of Computer Science and
Information Technology & Society, Vol. 02, No. 01, ISSN: 2249-9555, 2012

[2] Manish Gupta, B. Chandra and M. P. Gupta, “Crime Data Mining for Indian Police Information System”,

Computer Society of India, Vol. 40, No. 1, pp. 388- 397, 2008

[3] Kadhim B. Swadi Al-Janabi, “A Proposed Framework for Analyzing Crime Data Set Using Decision Tree and

Simple K-Means Mining Algorithms”, Journal of Kufa for Mathematics and Computer, Vol. 01, No. 03, pp. 08-24, 2011

[4] Shyam Varan Nath, “Crime Pattern Detection Using Data Mining”, IEEE Transactions on Knowledge and Data
Engineering, Vol. 18, No. 09, pp. 41-44, 2010

[5] Sasha Kapoor, Abhineet Kalra, “Data Mining for Crime Detection”, International Journal of Computer

Engineering and Applications, Volume VII, Issue III, September 14

[6] J.S. de Bruin, T.K. Cocx, W.A. Kosters, J.F.J. Laros, and J.N. Kok.Onto clustering criminal careers. In Proceedings of the ECML/PKDD2006 Workshop on Practical Data Mining: Applications, Experiencesand Challenges, pages 92–95, 2006

[8] Jyoti Aggarwal, Renuka Nagpal, Rajni Sehgal, “Crime Analysis using K-Means Clustering”,
International Journal of Computer Applications (0975 – 8887) Volume 83 – No4, December 2013.

[7] Nazlena Mohamad Ali1, Masnizah Mohd2, Hyowon Lee3, Alan F. Smeaton3, Fabio Crestani4 and Shahrul Azman Mohd Noah2 ,2010 Visual Interactive Malaysia Crime News Retrieval System.

[9] Sutapat Thirprungsri Rutgers University.USA ,2011 Cluster Analysis of Anomaly Detection in Accounting Data: An Audit Approach 1.

[10] Tony H. Grubesic, “On The Application of Fuzzy Clustering for Crime Hot Spot Detection”, Journal of Quantitative Criminology,Vol. 22, No. 1 (March 2006), pp. 77-105

[11] Tayal, D. K., Jain, A., Arora, S., Agarwal, S., Gupta, T., & Tyagi, N. Crime detection and criminal identification in India using data mining techniques. AI & SOCIETY, 30(1), 117-127.

[12] T. Wang, C. Rudin, D. Wagner, and R. Sevieri, Detecting Patterns of Crime with Series Finder, 2013.

[13] Qusay Bsoul, Juhana Salim, Lailatul Qadri Zakaria, “An Intelligent Document Clustering Approach to Detect Crime Patterns”, The 4th International Conference on Electrical Engineering and Informatics (ICEEI 2013)

[14] Aubaidan, Bashar, et al. "Comparative study of kmeans and k-means++ clustering algorithms on crime domain." Journal of Computer Science 10.7 (2014): 1197-1206.

[15] Mohammad Reza Keyvanpour, Mostafa Javideh, Mahammad Reza Ebrahimi, “Detecting and investigating crime by means of data mining: a general crime matching framework” Procedia Computer Science, vol. 03, pp872-880, 2011.

Received: 15 May 2019/Accepted 05 July 2019.