Research Article

Envision of Route Safety Direction Using Machine Learning

Aruna Pavate*, Anita Chaudhari and Rajesh Bansode

IT Department, Thakur College of Engineering, Mumbai University, India

*Corresponding Author: Aruna Pavate, IT Department, Thakur College of Engineering, Mumbai University, India.

Abstract

Background and Purpose: Despite of being lots of development women safety is always become one of the major and always under researched issue. As Delhi is proliferating become one of the most unsafe city in India. The number of women employment has decreased in Delhi over the few years because of insecurity. It is necessary to analyze the women safety in Delhi and provide the proper solution. There are many applications available which helps women not only to contact to the police, nearest person or friend but also each having their own benefits and limitations. To encourage people to travel in the night without any worry and can possible to contact the dearest one with single click with the help of this system.

Analysis Method: The system is designed using k-means clustering machine learning algorithm to improve the accuracy and reduce the access time. The idea is to predict the safe route and render the hindering measures for the women in order to confront the assertable crime situations and display the route with the rate of crime at the selected route and hence allows women to avoid certain places of special crimes. The dataset consists of 2115 instances that reported crimes across the Delhi city run through repository.

Findings: The system is built based on K- means clustering algorithm to get the accurate result. The proposed system trained using 1481 record and tested on 634 records. The designed system guarantees good results in-terms of finding the best the path based on the categories of the crime locations with the accuracy of the system with 79.7% with the time taken 70 ms.

Conclusion: The results of this system shows clustering of routes into different best route, better route, average, less secure route, lowest secured route. This work attempts to suggest the safe route with the type of crime happened on that route so that anyone can select the safest route from source in Delhi city.

Keywords: Women Safety; K-Means Algorithm; Clustering; Machine Learning; Crime; Safe Route

Introduction

Now a days, as the world has come closer because of advancement in technologies. Women are almost captured all the fields from sports, finance, health care, education and many more. In the Medieval period, women were taken into account as a major cause of destruction by major communities and don't give permission to the women go to outdoor or to get into the corporate and or social activities unlike men [1]. In the late middle ages women were involved in many fields. The statistics done by National Crime Records Bureau, the crime rate versus women registered around

4,037 incidences in the year 2000 in Chennai [2]. Due to increase in crime rates, the women growing persona is really a big challenge. Day to day, because of progressively insecure surroundings for Indian women working late at night is become very difficult. There are lots of very astonishing technologies available for women like video recording to encapsulate the cogent evidence of a crime, GPS tracker to racetrack people to your place etc. As its difficult for the tools to save you from the disaster but it assist you to send the alert to authorities or to your loved ones and to get ahead with assistance quicker whenever required. To promote the freedom not only

work but at late night, it is necessary to propose the system which helps to complete women security and safety. Such type of systems allows women to live without being afraid and work in the world with confidence, with informing their family, friends about their location. The purpose of this work is to design and develop system to help women safety based on machine learning concept. Table 1 shows the total number of crime rates versus women in India for ten different states during year 2013 to 2018 [3]. As per ob-

servations every year the rate of crime increases gradually. Figure 1 shows the crime Rates versus Women in the seven Major Cities in India during year 2005 - 06 and 2010 - 12. Table 2 shows the type of crime against women in Delhi city, India. In the present days Delhi has questionable about the safety of women. In Delhi women safety is the one of the biggest challenge despite the of the national capital of India.

Rank	State	2013	2014	2015	2016	2017	2018
1	Uttar Pradesh	75662.09	79874.18	84086.27	88298.36	92510.45	96722.55
2	Maharashtra	43434.42	44636.9	45839.38	47041.85	48244.33	49446.8
3	Andhra Pradesh	41971.47	43406.07	44840.67	46275.26	47709.86	49144.46
4	West Bengal	31828.74	33474.57	35120.41	36766.24	38412.07	40057.9
5	Madhya Pradesh	29155.17	29639.13	30123.09	30607.05	31091.01	31574.97
6	Gujarat	25225.33	26259.87	27294.41	28328.95	29363.49	30398.03
7	Odisha	16496.3	17360.02	18223.73	19087.44	19951.16	20814.87
8	Assam	17617.21	18176.91	18736.61	19296.31	19856.01	20415.71
9	Karnataka	16662.48	17341.5	18020.51	18699.52	19378.53	20057.54
10	Rajasthan	16719.21	17046.07	17372.92	17699.77	18026.62	18353.48

Table 1: Total number of crimes versus women in India for ten different state during year 2013 to 2018 [3].

Crime Category	2005 - 06	2008 - 09	2011 - 12
Rape	8.3	5.6	6.9
Kidnapping and abduction	14.0	16.3	22.9
Dowry deaths	1.6	1.5	1.4
Cruelty by husband and relatives	22.2	17.5	22.3
Molestation	9.7	7.3	7.9
Eve-teasing	2.5	1.7	2.2
Importation of girls	0.0	0.0	0.0
Total	58.2	49.8	63.6

Table 2: Crime category versus Women in one of the Major city Delhi [4].

Figure 1: Crime Rates versus Women in the Major Cities in India(2005–06 and 2010–12) [4].

To protect themselves people take different precautions. According to the [5] different parameters that are considering as safety measures as follows:
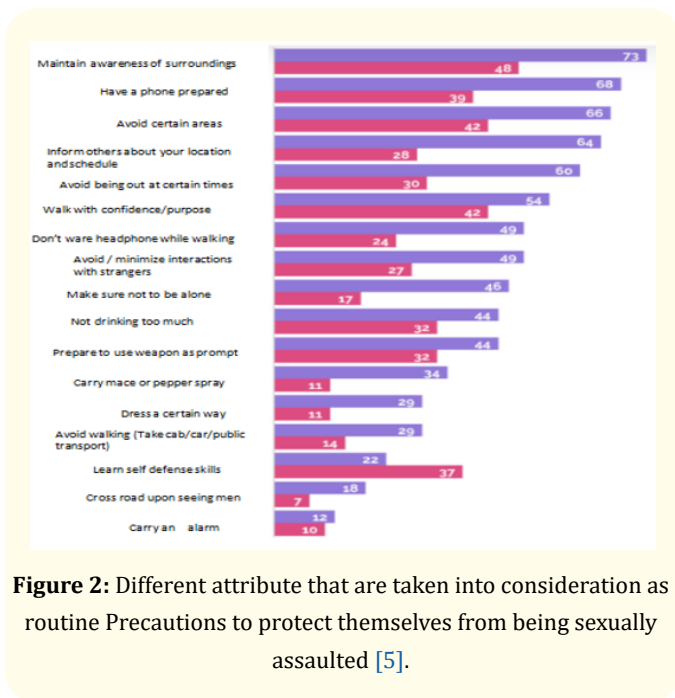

Figure 2: Different attribute that are taken into consideration as routine Precautions to protect themselves from being sexually assaulted [5].

Related work

There are lots of applocations available in this digital world, but theeach application having some limitations. Vith U App, Circle of 6 App,Life360 Family Locator App,I'm Shakti App, Famy Family Chat and Locator App,Nirbhaya: Be Fearless App,Watch Over Me App,Sentinel Personal Security SOS App,Secure Her App,Woman Safety Shield App, Spot Nsave Feel Secure App,Sprint Family Locator App [6]. Other tools like I Go Safely, Shake2Safety, Women Safety, Citizen Cop, Family, Locator - GPS Tracker, Eyewatch SOS for Women, My Safety Pin, Raksha, Watch. Me, bSafe, Himmat Plus etc [7].

Lawrence., *et al*. applied the concept of data mining and machine learning using weka tool. The analysis of crime data done using three different algorithm Additive Regression, Linear Regression and Decision Stump algorithms. The proposed system evaluated using different measurement metrics like Correlation Coefficient, Mean Absolute Error, Root Mean Squared Error, Relative Absolute Error and Root Relative Squared Error [8].

Sumanta Das., *et al*. proposed system based spatial clustering technique to predict the future crime location. SVM applied to decide the level of crime rate and K- means clustering algorithm used to cluster the data into hot spot class and cold spot class. The result were analysed for four different cities [9].

Md Abu Saleh proposed system using K-means algorithm to analyse and predict the crime in the town Chicago. The results were analysed and verified the flexibility and accuracy of the k-means algorithm [10].

All the above mentioned tools are developed to trace the location and send the message to the closest one about the safety. Its very important to know the safest route of journey. In this work, the system is proposed to demonstrate the safe path on condition of type of crime tumbled into that area.

Materials and Methods

Clustering is a machine clustering is the Logic used to in Machine learning programming to train the system based on previous data, learning method applied to group (cluster) the principle component that have similar characteristics/ data points. The outcome is that data points within the same group are more similar between them than other data points in other cluster [12]. There are two approaches are available like Hierarchical, Partition and further subdivided into different types [11,12]. The main aim of clustering algorithm is to obtain routes that are safe that means having similar data points and data points under different labels considered under unsafe group. As k- means clustering algorithm is almost suitable on all data types whether it is a figure, text or an image. Mean in the K-means refers to the average of all data points to find out the new centroid. The algorithm is very popular and you can easily find the implementation in almost all domains with study on different use cases. K- Means algorithm is most suitable for all types of data whether it is a figure, text, or an image. K-means is very flexible so it's an easy to apply on all types of data. K-means is an unsupervised learning method used to cluster similar data points. Unsupervised learning means there are no output labels to the points to learn the process to find out patterns. It is easy to implement and gives cost efficient performance. This algorithm is fast and deals with large amount of data.

Working of an algorithm

Dataset

The dataset consists of 2115 instances that reported crimes across the Delhi city and having attributes nm-pol, murder, rape, gang-rape, robbery, theft, assault, sexual-harassment, total-area, total-crime, longitude, latitude The dataset has details of five types of routes in the Delhi city of India. Each data point in that location represented by the Latitude and Longitudes. The main purpose of this work is to find the best safe route according to the crime rate analysis and the features that will be analysed into one of the cat-

egory as 0 for Best route, 1- Better route 2- Average 3-Less secure route, 4- Lowest secured route. Table 3 represents the sample of dataset.
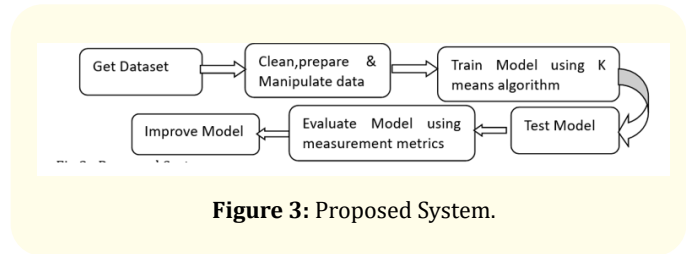

Figure 3: Proposed System.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	nm_pol	murder	rape	gangrape	robbery	theft	assualt mu	sexual har	totarea	totalcrime	long	lat	crime/area	area
2	CHITRANJAN PARK	2	6	1	35	442	19	7	2659330	512	77.2492	28.53632	192.5297	2.65933
3	DABRI	8	28	0	79	240	26	16	3401013	397	77.086	28.61268	116.7299	3.401013
4	MALVIYA NAGAR	3	28	1	33	694	63	15	1379854	837	77.20418	28.52989	606.5861	1.379854
5	CHANDNI MAHAL	1	8	1	23	529	19	7	5570696	588	77.23608	28.64361	105.5523	5.570696
6	MODEL TOWN	0	4	1	45	393	9	14	2689157	466	77.19369	28.70257	173.2885	2.689157
7	ANANDVIHAR	10	14	0	99	457	24	15	4558970	619	77.29373	28.65335	135.7763	4.55897
8	KASHMERE GATE	0	9	3	17	340	17	12	1627910	398	77.22618	28.66645	244.4853	1.62791
9	GOVIND PURI	1	4	3	31	692	5	5	8992712	741	77.26532	28.53067	82.40006	8.992712
10	BINDAPUR	4	4	1	64	417	11	8	2568080	509	77.06618	28.6091	198.2026	2.56808
11	NEW FRIENDS COLONY	6	7	1	47	609	14	10	4045387	694	77.2676	28.56234	171.5534	4.045387
12	SARITA VIHAR	1	12	0	12	211	8	1	7942774	245	77.30617	28.52299	30.84565	7.942774
13	TIMARPUR	8	17	1	60	626	23	5	6149745	740	77.2242	28.7072	120.3302	6.149745
14	KANJHAWALA	1	5	2	15	307	21	10	35490191	361	77.00361	28.72582	10.17182	35.49019
15	ANAND PARBAT	5	20	1	77	374	35	9	1948430	521	77.17376	28.66004	267.3947	1.94843
16	SAGARPUR	2	7	0	11	180	4	4	1283551	208	77.08644	28.59139	162.0504	1.283551
17	PRASHANT VIHAR	7	8	0	41	457	39	14	2076679	566	77.12757	28.71793	272.5506	2.076679
18	SOUTH CAMPUS	10	11	1	38	248	8	9	4518856	325	77.179	28.57051	71.92086	4.518856
19	ROHINI SOUTH	6	14	0	96	560	24	5	2649542	705	77.1159	28.70236	266.0837	2.649542
20	PUNJABI BAGH	9	24	0	64	1251	54	19	2577311	1421	77.13276	28.67441	551.3499	2.577311
21	DWARKA NORTH	3	2	0	10	104	10	13	30068548	142	77.01564	28.60369	4.722543	30.06855
22	BURARI	3	12	0	56	562	21	17	11333876	671	77.19735	28.73664	59.20305	11.33388
23	JAMA MASJID	4	19	0	13	317	10	1	3817050	364	77.23362	28.65284	95.3616	3.81705
24	ADRASH NAGAR	3	15	0	41	354	18	3	1646191	434	77.17532	28.70984	263.639	1.646191
25	CHANAKYAPURI	1	7	2	40	649	20	4	10993453	723	77.19606	28.60337	65.76642	10.99345
26	PANDAV NAGAR	5	16	3	27	351	23	10	1470613	435	77.30265	28.62038	295.795	1.470613
27	MIANWALI NAGAR	1	12	1	46	400	27	8	1729848	495	77.08879	28.67965	286.1523	1.729848
28	FATEHPUR BERI	7	10	0	60	683	7	9	35895595	776	77.15114	28.46549	21.61825	35.89559

Table 3: Sample dataset.

Proposed system

The proposed system implemented by applying K means algorithm. There are lots of applications available which helps to use in emergency from sending the alert messages to GPS tracker but every applications have their own limitations [6,7]. The proposed system not only provides the safe route but also allows to send the alert message as well as to call to the registered number. Safe route allows the user to look for the all possible safe route from the source to the destination and the user can select the safest route. Routes are suggested with different notations so the user can choose the route by themselves. The maps represented with different color, each color represent the particular type of crime rate happened in that area. Following are the steps applied to cluster the safe path from source to destination.

1. Depending on the value of K, K centroids are created randomly
2. The data point is to part of particular cluster if it is closer to that cluster's centroid than any other centroid. A centroid is the center of the data point. The centroid is calculated by using Euclidean distance [13]. Equation 1 shows calculation of centroid using Euclidean distance.

$$c(E) = \sum_{i=1}^{k} \sum_{j=1}^{ki} (\| xi - yi \|)^2 \quad (1)$$

Here ki is the number of data points in the cluster and k is the number of cluster centers.

1. To reduce the overall intra-cluster variation, the centroid is recalculated by taking the mean of all data points.
2. Step 2 and 3 repeated till the sum of distance between the data points and the corresponding centroid is minimized.

Results and Discussion

At the first we select a route randomly to find out the safety of the route at any given time of the day. It is necessary to rerun the algorithm with different starting centroid to provide best safe path, so the algorithm executed multiple times with considering various data points and evaluated using Kaufman approach. The selection of K gives the optimal solution. As shown in figure 4 the optimal value of K is 15. To create the optimal number of cluster, tried different number of clusters and calculated the sum of squared errors. The increase in value of K will cause decrease in the error sum while decrease will sharply increase the error sum. The point at which it defines the optimal number of clusters is refereed as 'elbow point' to find the best point for the value of K as shown in (Figure 5).

In this example, the elbow point is located in 15 clusters. The elbow point represents the optimal number of clusters and helps in representing the best pick for the value of K. and discussion must illustrate and interpret the reliable results of the study. As shown in figure 6 and 7 the red color symbols indicates the extreme crime rates on the route, moderate crime rates on the route showed by blue color symbol and safe path showed with green color symbol.
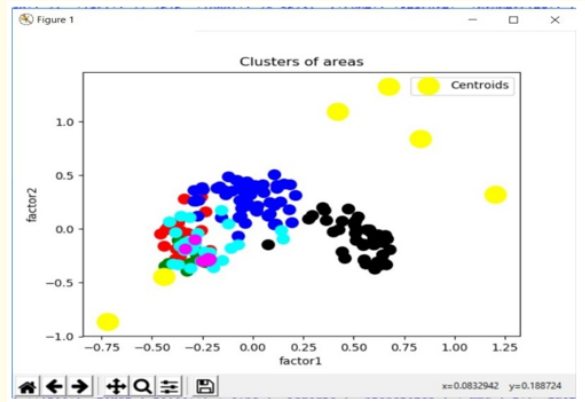

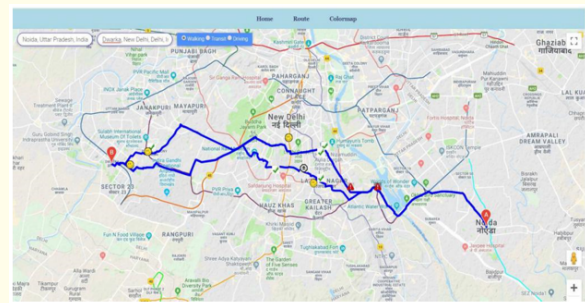**Figure 5:** Data points grouped into Five clusters based on Similarity.

Figure 6: Possible routes based on criteria of clustering from source to destination [Smart result].
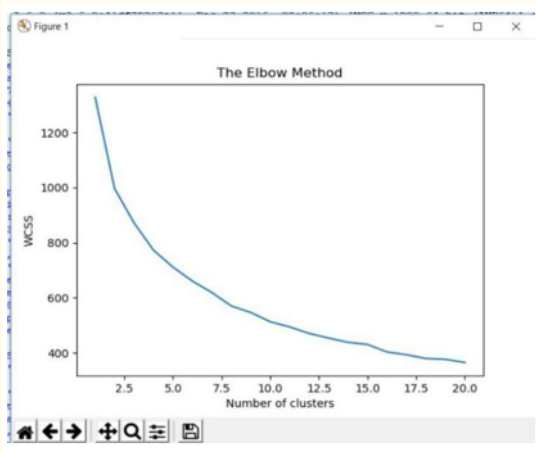

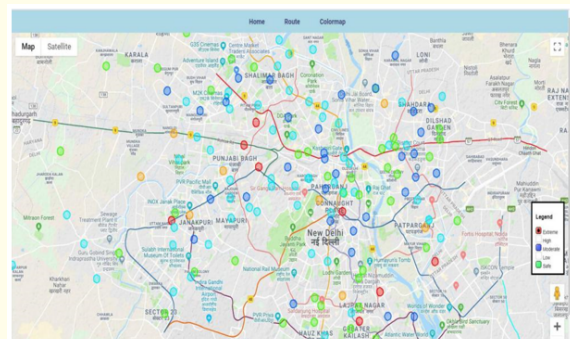**Figure 4:** Selection of Optimal number of Clusters.

Figure 7: Different colored symbols indicating the safety level. [Smart result].

Conclusion

K- means algorithm is having many benefits and can be used for almost any kind of data grouping as discussed in the previous section, but there are some factors where the performance of the system need to considered like selection of number of clusters, random selection of initial centroid, the generation of clusters with uniform sizes and spherical limitation if these modes handled carefully then the system behave spontaneously. The data in the real life is complex, noisy, and not organized well. Applying such algorithm in real life is very difficult because of clear conditions. The proposed system can be implemented using k-means clustering algorithm to provide the safest route.

Bibliography

1. Status of Womaninancient, Medievaland Modernperiod (2018).

2. Rashmi Rani Agnihotri HR and Malipatil KS. "A Brief Study on Women Problems in India". *International Journal of Development Research* 8.8 (2018): 22583-22587.

3. A crime map of India in R – Crimes against women (2019).

4. Shivani Satija and Amrita Datta. "Crime against Women and Children in Delhi Analysis of Secondary and Empirical Data". Economic and Political Weekly EPW (2015): 9.

5. https://today.yougov.com/topics/lifestyle/articlesreports/2019/03/28/women-safety-sexual-assault-awareness

6. https://www.hongkiat.com/blog/android-personal-safety-women-apps/

7. https://www.redbytes.in/best-women-safety-apps/

8. Lawrence Mc Clendon and Natarajan Meghanathan. "Using Machine Learning Algorithms to Analyze Crime Data". *Machine Learning and Applications: An International Journal (MLAIJ)* 2.1 (2015).

9. Sumanta Das and Malini Roy Choudhury. "A Geo-Statistical Approach for Crime hot spot Prediction". *International Journal of Criminology and Sociological Theory* 9.1 (2016): 1-11.

10. Md Abu Saleh and Ihtiram Raza Khan. "Crime Data Analysis in Python using K - Means Clustering". *International Journal for Research in Applied Science and Engineering Technology (IJRASET)* (2019): 2321-9653.

11. AK Jain MN Murty and PJ Flynn. "Data Clustering: A Review". in ACM Computing Surveys 31.3 (1999).

12. Pranav Nerurkar., *et al.* "Empirical Analysis of Data Clustering Algorithms". *Procedia Computer Science* 125 (2018): 770-779.

13. Gracy Joseph., *et al.* "Complete Animal Care System using Clustering Algorithm: CACS". *International Journal of Engineering Research and Technology* 5.1 (2017): 1-4.