International Journal of Computing and Digital Systems
ISSN (2210-142X)
Int. J. Com. Dig. Sys. 14, No.1 (Sep-2023)

http://dx.doi.org/10.12785/ijcds/140175

Identification of Location for Police Headquarters to Deal with Crime Against Women in India Using Clustering Based on K-Means Algorithm

Richa Indu[1], Sushil Chandra Dimri[1] and Bhawnesh Kumar[1]

[1]*Department of Computer Science and Engineering, Graphic Era Deemed to be University, Dehradun, INDIA*

Received 21 Mar. 2023, Revised 03 Aug. 2023, Accepted 10 Aug. 2023, Published 01 Sep. 2023

Abstract: During 2020-21, the rate of crime against women has considerably increased, especially the cases of domestic violence, kidnapping, and rape. Considering the fact that a woman better understands the problem of another woman, and to ensure women's safety in this era of technology, we proposed opening Police Headquarters with lady staff only. For this, we suggest the use of the latitude, longitude, and average weighted crime as clustering parameters for the K-means method to recognize the most probable location at which such headquarters can be opened. The parameter average weighted crime signifies the magnitude or intensity of nine different possible crimes against women and thus identifies the location where the maximum number of crimes against women are reported. Therefore, such locations are ideal for setting up the Lady Police Headquarters to further curtail any such crimes, and deliver swift protection as well as justice to the victim women. It is somehow similar to identifying clusters (locality) from clusters (blocks). Thence, using the suggested approach, the Indian law enforcing agencies including police can prevent and control crimes at the micro level and can also increase their reach.

Keywords: K Means Clustering, Police Headquarters, Crime against Women

1. Introduction

International Women's Day is celebrated every year on the 8th day of March, though, the brutality against women is increasing. Ensuring the safety of women is still a global concern today. In several places in India, feminine divines like Goddess Durga, Kali, Saraswati, and Laxmi are worshipped. Indian traditions and rituals symbolically treat girls and even women as these Goddesses. Strangely, criminality against women is intensifying in India. As per the (National Crime Records Bureau of India)NCRB's data, in contrast to 2020, the registered cases of crime targeting women in 2021 rises from 3.71 to 4.28 lakhs across India [1]. Moreover, the percentage of crimes targeting women per one lakh population also increased from 56% to 64.5% within the same year [1].

The Constitution of India provides certain fundamental rights to ensure equality, empowerment, and education among all. These rights are formed to oppose discrimination and dissolve the socio-economic, educational, and political difficulties faced by minorities, weaker sections of society, and especially 48.39% of the total population in India, i.e., women [2]. Despite several rights for women as well, the figures of crimes against women are not degrading. Moreover, even after efforts on central-, state government, and societal levels, the proliferation of certain crimes such as eve teasing, domestic violence, rape, dowry, child marriage, kidnapping, murder, and human trafficking are unbridled [1] [2].

In today's digital era, mobile phones are almost used by each individual. Thus, many researchers deployed sensors, Global Positioning Systems, Machine Learning, and other Artificial Intelligence tools to explore the potential of these evolving technologies in improving and ensuring women's safety. Recent efforts in this area include a weight-sensing device that easily fits in a palm, and on either squeezing or pressing send alert messages, and calls not only to the relatives of women but also to the nearest police stations [3]. Their idea is to minimize the reaction time so that crime will be stopped before it will be supposed to happen. Similarly, the work described in [4] used sound frequency sensors, which captured the pitch of female voices to evaluate the need for help. Moreover, these sensors are portable, do not consume much electricity, and can be implemented on any kind of cellular phone. Thus, overcomes the drawbacks of CCTV cameras such as cost, installation, electricity, and privacy. Another Smart Wearable Device is suggested in [5], which is based on Social Threatening Filter, and Dc-RFO-IoT. Once the emergency button of this device was pushed, its in-built application sends alerts to registered contacts and women's helplines, along with the minute-to-minute location. Moreover, with the implanted speaker, the gadget generates a high-frequency alarm alerting the nearby personnel too. In certain cases, the device also provides rescue suggestions. This device proved itself 93.43% accurate during testing with a safety ratio of 97.34

E-mail address: richaindu.cse@geu.ac.in, sushildimri.cse@geu.ac.in, bhawneshmca@gmail.com

L-1 and L-2 logistic regression (regularized) models by Dehingia and colleagues analyzed the help-seeking behavior of victims in an abusive relationship (or violence by the husband) [6]. The effort was to identify that during such circumstances, whether the victim tries to find help from anybody or formal sources like police, relatives, NGOs, etc. For this, they used the data obtained from a health survey conducted across India during 2015-16, where women from the age group of 15-49 years volunteered and answered questions related to their experiences of violence, empowerment, health, and socio-demographic characteristics. Kadam and Pasalkar performed sentiment analysis using tweets from Twitter and Instagram to understand the general mentality and emotions regarding women on social platforms [7]. Authors of both [8] and [9] reviewed the current technologies like IoT Devices, Machine Learning Techniques, Augmented Reality, Android Mobile Apps, and Artificial Intelligence used to prevent crime against women. Additionally, they also discussed the difficulties, restrictions, and future possibilities of these innovative methods to further enhance women's safety.

Aziz, Sharma, and Hussain used Decision Tree, Random Forest, Support Vector, Simple Linear, and Multiple Linear Regression methods to predict twenty-eight types of Indian Penal Code (IPC) recognizable crime from entire India [10]. The approach used data between 2014 to 2020 and provided year-, state-, and region-wise predictions, such that the least Mean Absolute Percentage Error of 0.2 was observed with Random Forest Regression. Similarly, NirmaleeKakati examined Panbazar(Assam)'s chronologies of Police Stations entirely dedicated to women, demonstrating a variety of issues encountered by them at work, and examined the interaction between women, gender roles, and police [11]. Their study revealed a sharp increase in the reported cases at the considered police station, which shows the probable signs of a rise in crimes against women and children, or more awareness of women. In another work, Amaral, Bhalotra, and Prakash analyzed that 22% more violence against women was reported on Women Lady Police Stations (WPS) in India, indicating an increase in reports associated with domestic violence, kidnappings, Self-reported intimate partner violence, and gender-specific mortality reports [12]. Further, their work also indicated the impact of police deterrence showed a slight improvement with a little decrease in the frequency of crime incidents. According to them, more crime reported at WPS shows an infusing belief of women in such official helps because WPS makes crime reporting less stigmatized, less corrupt, and more female-friendly.

It also highlights the fact that victims do not complain or register the case or reach out for official help due to the stigma of society, the high-profile culprit(s), gender biases at police stations, and lack of knowledge of rights, justification of crimes based on clothing, and victim blaming. So, addressing women-oriented crime by women themselves as also demonstrated in [12] and [13] showed a decrease in crime rate and an increase in women-oriented self-reporting of crime. Therefore, the proposed effort suggests the use of an unsupervised Kmeans Clustering approach to identify the most probable location at which Police headquarters with lady staff can be opened. For this latitude and longitude of the location are identified depending on the average weighted crime of that area (locality) of a block, which is equivalent to identifying clusters (locality) from clusters (various blocks). Hence, the proposed approach can overcome the aforementioned issues and allows more women to report injustice.

Further work can be organized into six Sections: The recent work related to the use of the Kmeans clustering algorithm is discussed in Section 2. Proposed Work is mentioned in Section 3 along with the classification of Crimes against women and different laws of the Indian Penal Code associated with that crime. Section 4 discusses Results on a dummy as well as the real dataset, followed by the Conclusion in Section 5 and Limitation in Section 6.

2. Related Work

A Quasi-Experimental design mentioned in [13] used Naive Bayes, Random Tree-, and Meta-Classifiers to categorize and predict crimes into five classes. 96.6% accuracy was the highest accuracy achieved with the meta-classifier. In contrast to supervised learning techniques, the unsupervised algorithms of machine learning can classify unlabeled data by grouping a collection of unlabeled objects. This process is known as clustering. Such a renowned clustering approach known to segregate a collection of observations into distinct groups (or clusters), is termed as K-Means technique. In general, it uses the sum of the squared distances between the observations and their assigned cluster mean is minimalized [14]. It means each observation is assigned to a cluster depending on the closest mean or centroid. Additionally, the optimal number of clusters (K) required to provide a clear view of the problem is decided either by the Elbow method or the Sillhoute method [15]. For an instance, if K=3, it designates the formation of three clusters. From 2011 to 2023, several pieces of research carried out utilizing K-Means Clustering to identify the location in the different scenarios are mentioned below.

Another KMeans clustering-based technique dividing the closest neighbors into groups based on how close they are to a mobile user was suggested in [16]. This nearest group to the mobile user is then utilized to determine the position of the mobile users. The evaluation findings show that the numbers of clusters, neighbors to be clustered, and the center points' beginning in the k-mean technique are all directly related to the efficiency of clustered KNN. Similarly, a solution to the location-allocation issue was suggested in [17], which used Stimulated Annealing with the K-means approach for a precise solution. Lee and Chung described the use of R for an effective criminal investigation and to provide a structured forensic data analysis using GPS and the hierarchical clustering model [18]. The work mentioned in [19] demonstrated a strategy for starting a new population and collecting higher-quality seeds by selecting the appropriate number of clusters automatically.In order to keep and manages a variety of gene sizes and shapes, the segmentation of niches

was performed using a density-based technique. The convergence to a local optimum can also be stopped using adaptive crossover and mutation probabilities. The best chromosomal centers are then retrieved and used as starting seeds in the K-means, which can adjust as necessary, to produce high-quality clustering. Experiments with four taxi GPS datasets manifested that NoeClust performs well and is successful at mining them for city context.

Likewise, Khalid and Herbert-Hansen utilized K-means clustering to create a novel framework that may be applied to International Location Decisions [20]. This approach produces a quantitative result devoid of individual subjectivity. Depending on the preferences of the decision-maker, the indicators can easily be added or removed. Thus, it was discovered that clustering with KMeans provide a quick and adaptable decision support framework. Authors of [21] deployed the density-first clustering tactic to create profiles comprising users' location information by placing a location service request to several neighboring users fulfilling the density metric criteria. After computing the analogous location profile among the service requester and the neighboring user, the collaborative filtering recommendation approach selects the best suitable serviceable location. Moslehi, Haeri, and Gholamian used K-means clustering-based algorithm to assign labels to each unknown cluster [22]. Furthermore, the work mentioned in [23], utilized the K-means clustering for substantiating the positioning of antennas to divide the testbed into zones, with 79% accuracy at K equals 4. Nigam used a semi-supervised method to solve the location prediction problem using both Hierarchical Density-Based Spatial Clustering of Applications with Noise (HDBSCAN, and Density-Based Spatial Clustering of Applications with Noise (DBSCAN) [24]. Apart from these, a geosocial dataset containing a user's latitude and longitude was also leveraged to identify the social groups of a user, and the distance between geosocial points was determined with the Haversine distance. The subsequent phase used Random forest classification in identified Points of Interest to determine the user's precise location, and also identify the services offered at the preferred point.

Additionally, Desineedi, Mahesh, and Ramadurai implemented one-step Markov chain modeling with KMeans to analyze the driving patterns clusters of intra-city buses in Chennai (India) [25]. For this, real-time GPS data had used throughout the peak and off-peak hours. With the computed probability transition matrix, they have obtained multiple candidate cycles whose durations range between 400 and 2800 seconds. The candidate cycle having the minimal average error was selected for concluding with three driving cycles, namely, morning-, off-, and evening peak hours. However, the evaluated driving cycles were found to be distinct from international as well as the existing ones. On the other hand, Fränti and S. Sieranoja investigated various ways to enhance the K-Means algorithm by modifying iterations and initialization [26]. To provide an insight into the current job market and ascertain the field demand among aspirants, the work discussed in [27] grouped different job interests. The types of employment, which are high or low in demand amongst aspirants were identified by the scatter plot portraying group density in clusters. Similarly, Wang and colleagues used an integrated evaluation model along with K-means clustering method to examine the location of private clinics in terms of profit gained [28]. Two versions of the K-Means algorithm were suggested by Eshete for an autonomous global distribution of web crawler containers to scrape specific URLs to the closest Google Cloud regions [29]. The first clustering algorithm takes into account the interval among URLs (websites) and regions of Google Cloud, and the latter approach considers both distance as well as cost. The experiential results inferred an optimization of 30% of the cost in price- and distance-based clustering. Soleman, Pramaita, and Sudarma used K-Means clustering to classify customer data into five clusters, namely, (i) a couple of minor transactions made by a large number of customers; (ii) several minor transactions entertained by a large number of clients; (iii) a handful of declining transactions involving a medium-sized clientele; (iv) a small number of significant transactions with a selective group of customers; and (v) an average amount of transactions with numerous customers [30].

Wu, Wei, Meng, Zhao, and Wang suggested a traffic density monitoring model, which takes the authentic location information from the vehicles themselves as well as maintains the privacy of the information [31]. To achieve this, the authors used a pseudonym and location anonymization servers, such that vehicles' location and identification were distinctly recorded. And to acquire location data for the anonymization servers, the probability distribution of historical locations was generated. Thus, creating anonymous yet authentic dummy location sets. In contrast with SimpMaxMinDistds and Metric Multi-Dimensional Scaling (MMDS) algorithms, the experimental results established their model as a better dummy location set generator, which was efficient in preserving location anonymity too. In another piece of work discussed in [32], noise algorithm-based K-Means clustering was used to aid road planning problems in urban areas. The main obstruction in handling such problems arises as difficulty in determining the accurate number of clusters and precisely recognizing their centroid. This problem could be effectively addressed with the noise algorithm, which adds noise judgment to improve the assignment of data points, and used silhouette coefficient, the sum of squares for error, Davies–Bouldin, and PBM-indexes [33] to analyze as well as enhance the outcome of clustering randomly. Results verified the superiority of this approach over fuzzy C-means, and K-means methods using GPS datasets from Aracaju, Beijing, Chongqing, Rome, and San Francisco. Additionally, Lin, Wu, and Pan used Fruit-Fly optimized K-means algorithm to boost distribution efficiency and the location solution for logistics distribution centers [34]. Tabianan, Velu, and Ravi used K-Means clustering to boost corporate profits and avoid customer churn based on shared behavioral traits [35]. For this, they worked on three clusters, namely, the event type, products, and categories, and assisted vendors by identifying the most profitable segments of the market by concentrating on the groups with comparable criteria. Again in [36], customer clusters were generated from the dataset of 200

customers, using K-means and agglomerative clustering algorithms.

Wickramasinghe and colleagues used wireless sensors to study high-resolution temperature measurements of a commercial building in terms of the health and performance of the HVAC (Heating, Ventilation, And Air Conditioning) controlling system [37]. After analyzing two such clusters, they computed the effectiveness of the existing space as well as the ideal number of clusters. Statistically, in this kind of problem, Time-Series clustering outperformed K-Means clustering. Likewise, keeping in mind the need for clusters with fixed centers to solve certain problems, two modified versions of the K-means algorithm, i.e., FC-K-Means, and FC-K-Means 2 were suggested in [38], where FC stands for Fixed Centers. As the name suggested, these methods locate the additional centroids by keeping the existent cluster centers fixed. According to the experimental findings, the FC-K-Means algorithm performed similarly to the original KMeans in terms of performance measures including Sum Squared Error, Silhouette, and Davies–Bouldin Indexes.

3. Proposed Work

A. Classification of Crimes Against Women

Any kind of physical or mental abuse or brutality to a woman is considered a crime targeting women. The United Nations(1993) described Violence against Women as - "an act of gender-based violence that results in or is likely to result in, physical, sexual, or psychological harm or suffering to women, including threats of such acts, coercion or arbitrary deprivation of liberty, whether occurring in public or private life" [39]. Further, according to the reports from NCRB and Indian Penal Code (IPC) [1], [40]–[42], the different crimes targeting women and the punishment associated with them respectively are as follows:

(i) *Domestic Violence* - Any kind of physical violence or misbehaving with a woman by her husband and his relatives is semantically treated as domestic violence. It includes beating, shoving, throwing objects, marital rape, etc. These acts of violence are punishable under Sec 498A of the Domestic Violence Act 2005, where a year of imprisonment and some fine are also imposed.

(ii) *Kidnapping* - Generally taking away any person without his/her consent is termed kidnapping. Similarly, abducting a woman or a girl for illegal trafficking, forced marriage, or to induce a physical relationship is an unlawful act that is punished under Sec 360, 361, and 366 of the IPC. The punishment for this crime is the incarceration of at most 7 to 10 years as well as fine.

(iii) *Eve Teasing, Stalking, and Insulting the modesty of women* - Following a woman or a girl in real and on social platforms, brushing and passing comments to harass her, and insulting her modesty (or embarrassing the womanhood publicly) is categorized under Stalking and Eve teasing. For both fines are imposed along with captivation of 3 to 5 years under Sec 509. Moreover, assaulting modesty is again a punishable offense under IPC Sec 354 and Sec 354B, with an imprisonment of 1 to 5, and 3 to 7 years respectively.

(iv) *Rape* - Rape refers to physical assault or Sexual Harassment of a woman or a girl. Unfortunately, it is the most common and prevailing crime against women, which law enforcement agencies and our societies failed to prevent. This crime is punishable under Section 376 and its sub-sections from A to D, such as Sec 376A (rape with murder), Sec 376B (marital rape), Sec 376C (rape by public servants), and Sec 376D (gang rape). Apart from a fine, such offenses have rigorous imprisonment for 7 to 20 years or life.

(v) *Honor Killing* - It is a series of an act of emotional and physical abuse, which ends in murder with the involvement of families, panchayats, or similar associations. The motive behind such killing in the name of honor of either family, caste, or community is to create terror and stop inter-caste, inter-religion, or love marriages. These actions are a blatant breach and misdemeanour of the fundamental rights provided to each Indian by the Constitution of India.

(vi) *Cyber Crimes* - With the advancement of technology, bullying, abusing, trolling, etc. are common. When such actions are performed with the intention to outrage the modesty of the women, then it is referred to as women-oriented cybercrimes. Under the Information Technology Act (2000), committing such crimes will lead to incarceration of a maximum of 3 years and a fine.

(vii) *Dowry* - The act of taking money, vehicles, land, or other valuable possessions in terms of marriage is still very common in India. Moreover, if such practices are opposed by the girl, or her family do not fulfill such demands of the boy's family, then the girl is ill-treated, beaten, attempted to kill, and even cold-bloodedly murdered by her husband or his relatives. However, such practice is punishable under the Dowry Prohibition Act of 1961 under IPC Sec 304B. According to this, the captivity of 7 years can be extended to life imprisonment in case of dowry death.

(viii) *Acid Attacks* - Throwing acid at a girl or woman to threaten or hurt her comes under acid attack. Nonetheless, the sale and purchase of acid without proper documents are banned in India, still, such evil is prevalent in society. Under Sec 326A and 326B of IPC, any serious damage due to an acid attack is punished with 7 years to life imprisonment with a fine.

(ix) *Women trafficking*– Women trafficking involves the selling and buying of girls and/or women, who are illegally imported or exported, either from India or from a foreign country for exploitation, forced marriage, domestic enslavement, organ trade, surrogacy, forced labor, prostitution, and other illegal activities. Such offenses are punishable by 3 years to life incarceration and a fine under Sec.370, 372,

and 373.

Based on the aforementioned study and the records of NCRB of crimes against women in 2019 [42], the following nine basic Crime Heads (CH) are identified. Each crime head comprises sets of different crimes punished under the same sections. These are mentioned in Table I.

TABLE I. 9-CRIME-HEADS AND DIFFERENT CRIMES UNDER EACH HEAD

Crime Heads	Different Crimes
CH1	Rape, Attempt to Rape, Rape with Murder, Gang Rape
CH2	Dowry, Dowry deaths, Abusing for Dowry
CH3	Abetment to Suicide
CH4	Acid Attack, attempt to acid attack
CH5	Kidnapping and Abduction, Kidnapping and compelling her to marriage
CH6	Cruelty by Husband and his relatives, Domestic Violence
CH7	Human and Immoral Trafficking, Importation of girls from foreign countries, Procuration and selling of minor girls
CH8	Assaulting women for outraging her modesty, Stalking and Eve teasing, Insulting the modesty of women,
CH9	Cyber-crimes targeting women including Publishing or Transmitting of Adult Content

The distribution of crimes in India according to the NCRB reports for the year 2019 [42] is portrayed in Figure 1 as per the 9-Crime Heads.

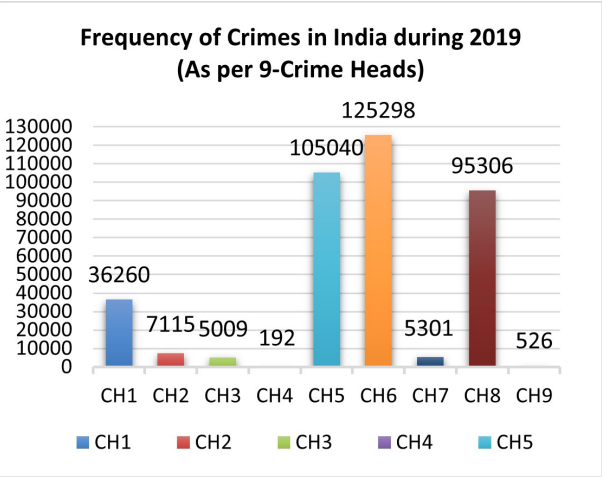

Figure 1. Bar graph representing the number of crimes under each crime head in entire India during 2019 from the NCRB records [42].

B. Proposed Algorithm

The proposed concept not only used the concept of weights but also the latitude and longitude to ensure better location selection for opening the Police Headquarters with lady staff. These Headquarters (HQ) will solve the women-oriented crime cases in the nearby localities, sub-blocks, and blocks. It also helps in safeguarding women's rights and will make them feel safe.

With the concept of weights, each crime head is assigned a weight depending on the severity of the crime and IPC. These weights are mentioned in Table II, which are then used to compute the weighted crime for a block (bigger area) using Eq. 1.

$$w_{c_i} = \frac{w_i x_i}{\sum w_i} \tag{1}$$

where, $\sum w_i = 0.1+0.2+0.3+\ldots+0.8+0.9 = 4.5$, from $i = 1$ to 9, w_i is the weight of the i^{th} crime head, and x_i is the number of cases reported under i^{th} crime head.

TABLE II. DIFFERENT CRIME HEADS IN DECREASING ORDER OF WEIGHTS

S.No.	Crime Head	Weightage
1	CH1:Rape	0.9
2	CH3: Abetment to Suicide	0.8
3	CH7: Human and Immoral Trafficking	0.7
4	CH4: Acid Attack	0.6
5	CH5: Kidnapping and Abduction	0.5
6	CH6: Cruelty by relatives	0.4
7	CH8: Insulting Modesty	0.3
8	CH2: Dowry	0.2
9	CH9: Cybercrimes	0.1

Once the weighted crime for a macro-level is calculated, i.e., a district, further an average weighted crime at the block is computed, comprising all the localities within that block. Assuming that there exists a total of p blocks, the average weighted crime for each locality of the p blocks can be evaluated as:

$$W_{c_{ij}} = \frac{1}{c_n} \sum_{j=1}^{n} w_{c_j} \tag{2}$$

Here, j denotes the localities in i^{th} block, while the maximum localities in the i^{th} block is represented by n. The c_n is the number of j-localities in the i-blocks, and w_{c_j} is the weighted crime in the j^{th} locality of the i^{th} block (computed using Eq. 1).Thus, forming the clusters, comprising localities of various blocks, where the latitude (on the x-axis), longitude (on the y-axis), and average weighted crime of the j^{th} locality of the i^{th} block (on the z-axis) is envisioned.

Now to achieve the objective of identifying the most probable locations for police HQ, the KMeans clustering is used. It divides the existing p-clusters into more manageable K-clusters to identify the hidden pattern in the data by clubbing alike observations. Using the two fundamental properties, first, all the observations within a cluster should be as analogous as possible and observations from different clusters should be as distinct as possible, here, the K-means clustering forms clusters of clusters. Thus, to identify the value of (K), i.e., the

optimal amount of clusters from p-clusters, the elbow point technique is utilized. This method uses a plot between the values of K (on the x-axis) and inertia (the sum of the squared distance from each observation to the centroid on the y-axis). The same is also shown in Figure 4, so as the curve reaches an elbow point or depicts a linear decline then that respective value of K is selected as the required number of clusters to be formed.

Thereafter, K-centroids are randomly initialized to find the similarity within the unlabeled observations and to cluster them together. Generally, the distance metric used for calculating inertia is the Euclidean distance (for numerical features), and Manhattan distance (for categorical features). Since we have to handle the numerical data, which is the latitude, longitude, and average weighted crime, thus we use Euclidean distance to measure the distance between the observations and the centroids. Mathematically, for a set of cluster centroids $\{C_1, C_2, \ldots, C_K\}$ with a maximum of K centroids; C_1 denotes the centroid of cluster-1, C_2 represents the centroid of cluster-2, and so forth up to C_K, which stands for the centroid of the K^{th} cluster. The Euclidean distance (d) in the three-dimensional plane is given by:

$$d = \sqrt{(y_1 - c_1)^2 + (y_2 - c_2)^2 + (y_3 - c_3)^2} \qquad (3)$$

Here, (c_1, c_2, c_3) are the coordinates of a centroid C_m from the set of centroids, where $m \leq k$, and $Y(y_1, y_2, y_3)$ are the coordinates of the observation (or data point) from the total sample size. Now, the observations are assigned to the cluster which possesses a minimal distance from the centroid of that cluster. Based on the minimal intra-cluster Euclidean distances, once all the observations are assigned to the respective clusters, the coordinates of centroids are updated. For this, the average of all the observations in cluster C_m is computed using Eq. 4.

$$NewCentroid = \frac{1}{n} \sum Y_n \qquad (4)$$

where n is the total observations in cluster C_m, and $\sum Y_n$ is the sum of the n-observations in m^{th} cluster. This average of all the observations belonging to a cluster yields the coordinates of the new centroid of that cluster. The entire process of updating the centroid and assigning observations to the respective clusters based on the minimum Euclidean distance is carried out iteratively until the coordinates of centroids stop changing very much. In this way, the K-clusters can be formed out of p-clusters, where K is selected using the elbow point method. The centroids of these K-clusters include the latitude, longitude, and average weighted crime as per j-localities of p-blocks, and thus provide an ideal location for opening the Lady police HQ. The proposed method can be summarized in Figure 2.

Algorithm to identify the probable locations for Lady Police HQ

Input: *Latitude, Longitude, and Frequency of crimes under different crime heads for p blocks.*

Output: *K - Centroids yielding optimal* **[Latitude, Longitude, Average Weighted Crime]** *for opening Police HQ with Lady staff.*

Step 1. Assign weights depending on the severity of the crime using Table II, and compute weighted crime for each of the p blocks with Eq. (1).
Step 2. Evaluate the average weighted crime for each of the j localities in the i^{th} block using Eq. (2).
Step 3. Using the elbow point method, identify the optimal number of clusters, i.e., K from p-blocks.
Step 4. Initialize the coordinates for K centroids randomly.
Step 5. Repeat Until Eq. (4) stops changing significantly:
For p-blocks indicating p-clusters:
(i) Compute Euclidean distance between the C_m centroid and observations using Eq. (3), such that $m \leq k$.
(ii) Assign a respective cluster to each observation Y depending on the minimal intra-cluster distance between Y and C_m.
(iii) Update the coordinates of the centroids using Eq. (4)

Figure 2. Proposed methods' pseudo-code

4. Results and Discussion

The proposed approach is tested on randomly created dummy data and real data from NCRB [42].

A. Results on Pseudo data

Let us consider the latitude and longitude of the pseudo location would range between 28.4400 to 31.28000°N and 77.3500 to 81.01000°E. Then, the number of crimes under each aforementioned crime head is generated, which varies between 0 to 20. Again, we assumed that a total of 13 blocks are there in that span of longitude and latitude (probably in a district of the State), where the random number of localities (including sub-blocks) is generated for each block, such that a block must have a minimum of 10, and a maximum of 120 localities. The distribution of localities in each block is as follows:

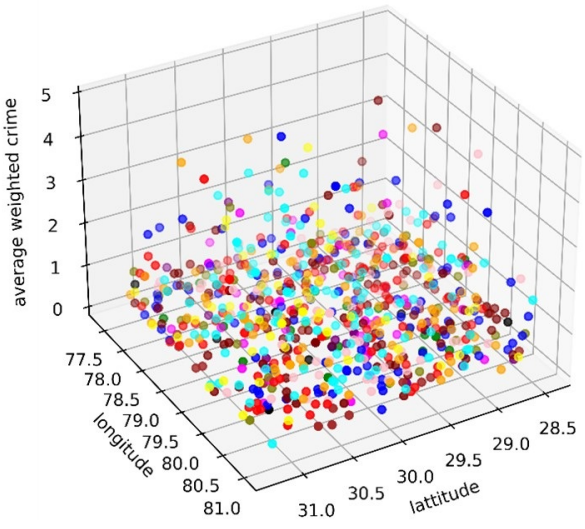

Figure 3. 13-colored points representing 13 blocks of a district

There exist 110, 20, 102, 10, 91, 72, 40, 42, 43, 79, 49, 21, and 111 randomly generated localities in block i to n, where, i begins from 1 to n, and n is the maximum number of blocks in that district (i.e., 13 as assumed). In this way, the distribution of a total of 790 localities of 13 blocks of one district is depicted in Figure 3.Now, we have 13

distinct clusters each representing one of the blocks. Now, from these p-clusters (or 13-clusters) of blocks, we have to identify the optimal amount of K clusters. Thus, Elbow point tactic is used (as denoted in Figure 4) for creating K-clusters from initial p-clusters.

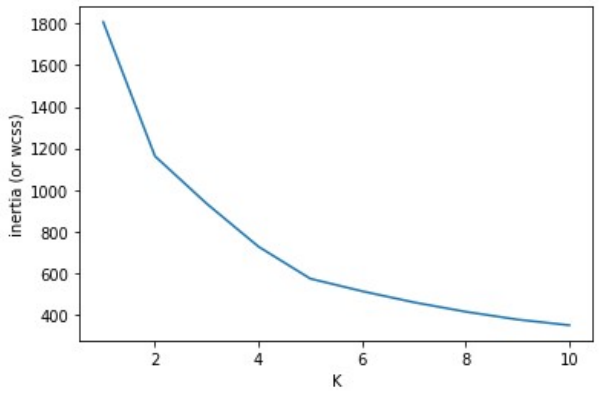

Figure 4. Elbow point curve representing the optimal value of K equal to 5

Consequently, after implementing the K-means Clustering Algorithm, the final five clusters of 13 clusters are portrayed in Figure 5. Additionally, the probable locations of the Police Headquarters could be at the latitudes and longitudes mentioned in Table III, along with the average weighted crime of that location.

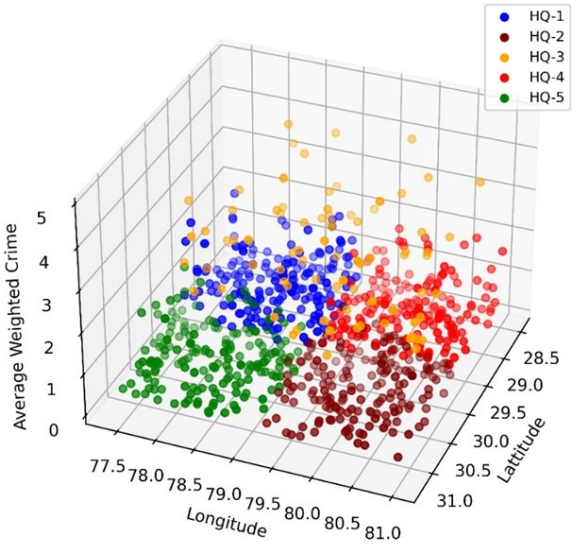

Figure 5. The final K=5 clusters on the dummy dataset

TABLE III. PROBABLE LOCATIONS OF POLICE HQ WITH LADY STAFF (ON PSEUDO DATA)

HQ	Latitude (°N)	Longitude (°E)	Average weighted crime
1	29.16412004	78.33330168	0.73556211
2	30.44363288	80.12348522	0.6515164
3	29.95844423	79.46041188	2.90049963
4	29.08116097	80.06708591	0.69042943
5	30.58952536	78.25238824	0.73246068

B. *Results on Real Dataset*

Similarly, the proposed approach is tested on the real dataset obtained from NCRB reports of 2019 [42]. It comprises 76 districts of Uttar Pradesh (UP) with different crime heads. The actual latitude and longitude are added to the dataset, whereas the different available frequencies of crimes are adjusted according to proposed crime heads. Now, the suggested approach is used in UP state and its 76 districts, where Eq. 1 assigns and evaluates the weighted crime as per each crime head and frequency of crime. Similarly, Eq. 2 computes the average weighted crime for each district, where the number of blocks in each of the 76 districts is randomly generated using Python programming. In this way, the ideal locations for women-oriented Police HQ are [26.53110°N, 80.04650°E] for HQ-1 with an average weighted crime of 1.0413, [28.46820°N, 78.19270°E] for HQ-2 with an average weighted crime of 1.1695, and [26.24930°N, 82.72630°E] for HQ-3 with an average weighted crime of 0.8231. The same is also mentioned in Table IV along with the final formation of K-clusters depicted in Figure 6.

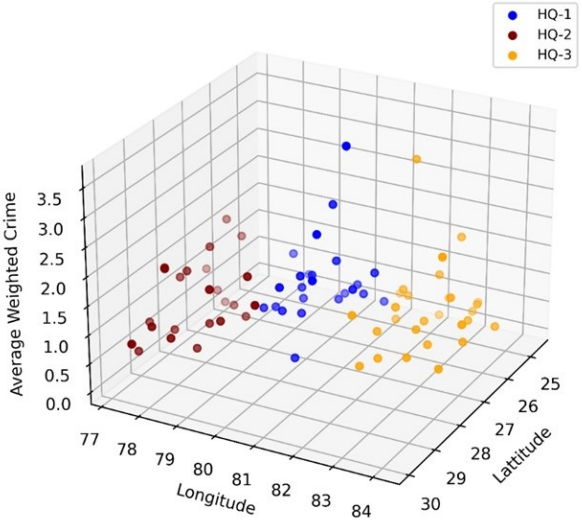

Figure 6. 3 clusters for UP state with 76 districts

TABLE IV. PROBABLE LOCATIONS OF LADY POLICE HQ ON A REAL DATASET COMPRISING 76 DISTRICTS OF UP

HQ	Latitude (°N)	Longitude (°E)	Average weighted crime
1	26.5311	80.0465	1.0413
2	28.4682	78.1927	1.1695
3	26.2493	82.7263	0.8231

5. Conclusion

A woman fulfills many responsibilities, from being a daughter, sister, wife, mother, and so forth. She deserves to be respected and treated equally. Providing and ensuring rights is not limited to or scope of clothes or family matters, but their safety and role in the development of a healthy society is a must and cannot be overlooked. Therefore, we used KMeans clustering to form clusters of localities from the clusters of blocks of a district to suggest a location in terms of latitude, longitude, and average weighted crime in that locality of the block for opening the police HQ. Since these police HQ will have lady staff only and are specially designed to address crime against women, this approach will ensure more reporting of crime and quick measures for readdressing their problems. Moreover, the suggested effort will avoid the stigma of gender biasedness, which is another cause of low reporting of women-oriented crimes in normal Police stations. Finally, the opening of such a police HQ not only enhances women's safety but also their belief in the official system to get justice.

6. Limitation

To measure the worth of work for a large amount of data, we have to use a pseudo or dummy data set due to the lack of availability of block-level data.

References

[1] G. Rajmohan, "Statistical data of violence against women and their legal status in india," *Legal Service India E-Journal*. [Online]. Available: https://www.legalserviceindia.com/legal/article-4558-statistical-data-of-violence-against-women-and-their-legal-status-in-india.html

[2] T. W. B. Group, "World bank staff, population, female (% of total population) - india." [Online]. Available: https://data.worldbank.org/indicator/SP.POP.TOTL.FE.ZS?locations=IN

[3] N. R. Wagh and S. R. Sutar, *An Enhanced Security of Women and Children Using Machine Learning and Data Mining Techniques*. Wiley, jan 2022, pp. 423–446.

[4] R. H. Sagar, L. K. S. R. Goud, A. Sharma, T. Ashraf, and A. P. Agrawal, *Women Assault Detection and Providing Help Using Pitch*. Cham: Springer Nature Switzerland, 2022, vol. 1738, pp. 531–543.

[5] K. Srinivas Rao, D. V. Divakara Rao, I. Patel, K. Saikumar, and D. Vijendra Babu, "Automatic prediction and identification of smart women safety wearable device using dc-rfo-iot," *Journal of Information Technology Management*, vol. 15, no. Special Issue, pp. 34–51, 2023. [Online]. Available: https://jitm.ut.ac.ir/article_89410.html

[6] N. Dehingia, A. K. Dey, L. McDougal, J. McAuley, A. Singh, and A. Raj, "Help seeking behavior by women experiencing intimate partner violence in india: A machine learning approach to identifying risk factors," *PLOS ONE*, vol. 17, no. 2, p. e0262538, feb 2022.

[7] R. Kadam and J. C. Pasalkar, "Analysis of women safety using machine learning on tweets," *International Journal of Research Publication and Reviews*, vol. 3, no. 11, pp. 2576–2579, nov 2022. [Online]. Available: https://ijrpr.com/uploads/V3ISSUE11/IJRPR8145.pdf

[8] P. Lonbale, P. Mane, S. Kharade, and B. J. Mohite, "Role of it in women safety," *International Research Journal of Modernization in Engineering Technology and Science*, vol. 5, no. 1, 2023. [Online]. Available: https://www.irjmets.com/uploadedfiles/paper/issue_1_january_2023/33221/final/fin_irjmets1674988774.pdf

[9] K. Ashok, A. B. Gurulakshmi, M. B. Prakash, R. Poornima, N. S. Sneha, and V. Gowtham, "A survey on design and application approaches in women-safety systems," in *2022 8th International Conference on Advanced Computing and Communication Systems (ICACCS)*. Coimbatore, India: IEEE, mar 2022, pp. 101–110.

[10] R. M. Aziz, P. Sharma, and A. Hussain, "Machine learning algorithms for crime prediction under indian penal code," *Annals of Data Science*, jul 2022.

[11] NirmaleeKakati, "Women in police service: A case study of all women police station, assam," *International Journal of Research in Social Sciences*, vol. 9, no. 4, pp. 720–730, 2019. [Online]. Available: https://www.indianjournals.com/ijor.aspx?target=ijor:ijrss&volume=9&issue=4&article=042

[12] S. Amaral, S. Bhalotra, and N. Prakash, "Gender, crime and punishment: Evidence from women police stations in india," *SSRN Electronic Journal*, no. 9002, 2021.

[13] S. S. Kshatri, Devanand, Bhonsle, R. Verma, and A. G. Pillai, "Crime detection approach using big data analytics and machine learning," *NeuroQuantology*, vol. 20, no. 8, pp. 1480–1495, may 2022.

[14] I. H. Sarker, "Machine learning: Algorithms, real-world applications and research directions," *SN Computer Science*, vol. 2, no. 3, mar 2021.

[15] S. Tripathi, A. Bhardwaj, and P. E, "Approaches to clustering in customer segmentation," *International Journal of Engineering & Technology*, vol. 7, no. 3.12, p. 802, jul 2018.

[16] B. Altintas and T. Serif, "Improving rss-based indoor positioning algorithm via k-means clustering," in *17th European Wireless 2011 - Sustainable Wireless Technologies*, Vienna, Austria, 2011, pp. 681–685.

[17] M. Mohammadkhanloo and M. Bashiri, "A clustering based location-allocation problem considering transportation costs and statistical properties," *International Journal of Engineering*, vol. 26, no. 6 (C), pp. 597–604, jun 2013.

[18] C. Lee and M. Chung, "Digital forensic for location information using hierarchical clustering and k-means algorithm," *Journal of Korea Multimedia Society*, vol. 19, no. 1, pp. 30–40, 2016.

[19] X. Zhou, J. Gu, S. Shen, H. Ma, F. Miao, H. Zhang, and H. Gong, "An automatic k-means clustering algorithm of GPS data combining a novel niche genetic algorithm with noise and density," *ISPRS International Journal of Geo-Information*, vol. 6, no. 12, p. 392, dec 2017.

[20] W. Khalid and Z. N. E. Herbert-Hansen, "Using k-means clustering in international location decision," *Journal of Global*

Operations and Strategic Sourcing, vol. 11, no. 3, pp. 274–300, 8 2018.

[21] P. Wang, J. Yang, and J. Zhang, "A strategy of cluster-based distributed location service," *Mobile Information Systems*, vol. 2019, pp. 1–13, 5 2019.

[22] F. Moslehi, A. Haeri, and M. R. Gholamian, "A novel selective clustering framework for appropriate labeling of the clusters based on k-means algorithm," *Scientia Iranica*, vol. 27, no. 5, pp. 2621–2634, 2020.

[23] K. F. P. Wye, E. Kanagaraj, S. M. M. S. Zakaria, L. M. Kamarudin, A. Zakaria, and K. K. N. Ahmad, Eds., *RSSI-based Localization Zoning using K-Mean Clustering*, Pulau Pinang, Malaysia, 2019, vol. 705.

[24] S. Nigam, "Predicting next location and recommend service using geo-social data," *Journal of Emerging Technologies and Innovative Research*, vol. 6, no. 6, pp. 446–451, 2019. [Online]. Available: https://www.jetir.org/papers/JETIR1907524.pdf

[25] R. M. Desineedi, S. Mahesh, and G. Ramadurai, "Developing driving cycles using k-means clustering and determining their optimal duration," *Transportation Research Procedia*, vol. 48, pp. 2083–2095, 2020.

[26] P. Fränti and S. Sieranoja, "How much can k-means be improved by using better initialization and repeats?" *Pattern Recognition*, vol. 93, pp. 95–112, 9 2019.

[27] F. J. M. Shamrat, Z. Tasnim, I. Mahmud, N. Jahan, and N. I. Nobel, "Application of k-means clustering algorithm to determine the density of demand of different kinds of jobs," *International Journal Of Scientific & Technology Research*, vol. 9, no. 2, pp. 2550–2557, 2020. [Online]. Available: https://www.ijstr.org/final-print/feb2020/Application-Of-K-means-Clustering-Algorithm-To-Determine-The-Density-Of-Demand-Of-Different-Kinds-Of-Jobs.pdf

[28] X. Wang, C. Shao, S. Xu, S. Zhang, W. Xu, and Y. Guan, "Study on the location of private clinics based on k-means clustering method and an integrated evaluation model," *IEEE Access*, vol. 8, pp. 23 069–23 081, 2020.

[29] G. A. Eshete, "Autonomous global distribution of container workload using k-means clustering algorithm," pp. 1–75, 2020. [Online]. Available: https://www.duo.uio.no/bitstream/handle/10852/81211/1/Getinet_NSA_thesis_2020.pdf

[30] C. D. O. Soleman, N. Pramaita, and M. Sudarm, "Classification of loyality customer using k-means clustering, studicase : Pt. sucofindo (persero) denpasar branch," *International Journal of Engineering and Emerging Technology*, vol. 5, no. 2, pp. 160–167, 2020.

[31] L. Wu, X. Wei, L. Meng, S. Zhao, and H. Wang, "Privacy-preserving location-based traffic density monitoring," *Connection Science*, vol. 34, no. 1, pp. 874–894, 2022.

[32] X. Ran, X. Zhou, M. Lei, W. Tepsan, and W. Deng, "A novel k-means clustering algorithm with a noise algorithm for capturing urban hotspots," *Applied Sciences*, vol. 11, no. 23, p. 11202, 11 2021.

[33] M. K. Pakhira, S. Bandyopadhyay, and U. Maulik, "Validity index for crisp and fuzzy clusters," *Pattern Recognition*, vol. 37, no. 3, pp. 487–501, mar 2004.

[34] T. X. Lin, Z. H. Wu, and W. T. Pan, "Optimal location of logistics distribution centres with swarm intelligent clustering algorithms," *PLOS ONE*, vol. 17, no. 8, p. e0271928, 8 2022.

[35] K. Tabianan, S. Velu, and V. Ravi, "K-means clustering approach for intelligent customer segmentation using customer purchase behavior data," *Sustainability*, vol. 14, no. 12, p. 7243, 6 2022.

[36] H. Lone and P. Warale, "Cluster analysis: Application of k-means and agglomerative clustering for customer segmentation," *Journal of Positive School Psychology*, vol. 6, no. 5, pp. 7798—7804, 2022. [Online]. Available: https://journalppw.com/index.php/jpsp/article/view/8854/5773

[37] A. Wickramasinghe, S. Muthukumarana, D. Loewen, and M. Schaubroeck, "Temperature clusters in commercial buildings using k-means and time series clustering," *Energy Informatics*, vol. 5, no. 1, 2 2022.

[38] M. Ay, L. Özbakır, S. Kulluk, B. Gülmez, G. Öztürk, and S. Özer, "FC-kmeans: Fixed-centered k-means algorithm," *Expert Systems with Applications*, vol. 211, no. 3, p. 118656, jan 2023.

[39] L. Sabha. [Online]. Available: https://loksabhadocs.nic.in/Refinput/New_Reference_Notes/English/Crimeagainstwomen.pdf

[40] NCRB, *Crime Against Women, Crime in India*, 2015, ch. 5, pp. 81–92. [Online]. Available: https://ncrb.gov.in/sites/default/files/crime_in_india_table_additional_table_chapter_reports/Chapter%205-15.11.16_2015.pdf

[41] A. Bajaj, "Crimes against women - a legal perspective," 2023. [Online]. Available: https://www.indianbarassociation.org/crimes-against-women-a-legal-perspective/

[42] NCRB, "District-wise crime against women [data]," 2019. [Online]. Available: https://ncrb.gov.in/en/search/node/crime%20against%20women

Richa Indu currently pursuing Ph.D. in Computer Science and Engineering from Graphic Era Deemed to be University. She accomplished M.Tech(Hons) in Computer Science and Engineering from Uttarakhand Technical University, Dehradun and M.Sc(Gold medalist) in Information Technology from Hemavati Nandan Bahuguna University, Srinagar (Garhwal), INDIA. She has published more than four papers in conferences and journal. Her research interest includes machine learning, programming languages, data sciences and designing algorithms.

Prof. Sushil Chandra Dimri currently serving Graphic Era Deemed to be university as a professor in CSE Department. He received M.Tech from IIT Dhanbad and Ph.D in Computer Science from Kumaon University, Nainital, Uttarakhand, India. He has 22 years of experience in teaching of UG and PG level degree courses. Authors of many books and published more than 60 papers in national/international conferences and journals. His areas of interest are Algorithm design, Resource optimization , Machine Learning and Computer Graphics.

Bhawnesh Kumar , Assistant Professor, Department of Computer Science and Engineering in Graphic Era Deemed to be University, Dehradun. He had received MCA from Kurukshetra University, Kurukshetra, Haryana, India, M.Tech (CSE) from R. G. T. U Bhopal, M.P (State University), India. He has more than 16 years of experience in teaching of UG and PG level degree courses as a Lecturer/Assistant Professor in various academic organizations. He has published more than 20 research papers in reputable international journal and conferences. His areas of interest are agile software development, wireless sensor network, computer graphics and machine learning.