

Identifying the Optimal Location to Set up an Emergency Room Using Machine Learning Algorithms

DR.G.SIVAKAMASUNDARI,

Department of Computer Science and Engineering, National Engineering college,Kovilpatti.

MS.B.SHUNMUGAPRIYA,

Department of Computer Science and Engineering, National Engineering college,Kovilpatti.

MR.A.SHENBAGHARAMAN,

Department of Computer Science and Engineering, National Engineering college,Kovilpatti.

MR.K.RAJKUMAR,

Department of Computer Science and Engineering, National Engineering college,Kovilpatti.

Abstract: *As per the the Allstate Canada Safe Driving Study report, Toronto was in 69th position in accident rate among Canadian cities with an average of 6.45 accidents per 100 cars in 2014-2015. North York and Ajax, the borders of Toronto the accident rate was even worse with an average of 7.02 to 7.12 per 100 cars [1]. The higher the accident rate, the higher the death rate. By reducing the time lag between the accident and the initiation of medical care, one can prevent death or permanent disability. The distance between the accidents zones the emergency room play the vital role in reducing the death rate due to accident. As per the report, most of the accidents were at the outskirts of the city rather than within the city. But usually the most of the emergency rooms are within the city. In such cases mostly, the emergency rooms were far away from the accident zones. The objective of the work is to predict the most suitable place for establishing the emergency rooms using machine learning algorithms. Accident zones in Toronto, the dataset was taken from Toronto public service data portal and locations of emergency rooms were retrieved from Foursquare API. After mapping both the data set, the accident zones near the emergency rooms (which are at the distance of 1 km) are removed. Then accident dense area was found using hierarchical dbscan. K nearest neighbor algorithm is used to address the outliers. The suitable (core) location for the emergency room was found by taking the mean of each cluster. The distance between the core location and the emergency room was found. The core location with the longest distance was considered as the best place for establishing the new emergency room.*

Keywords: *Machine learning. Clustering, Hierarchical dbscan algorithm, Classification, k Nearest Neighbor algorithm.*

Introduction:

Nowadays death rate due to accidents are increasing day by day and road accidents are estimated to be the eighth leading cause of death globally for all age groups, especially for children and young people 5–29 years of age. Each year approximately 1.35 million people are killed on road accidents. Every day, almost 3,700 people are killed globally in road accidents involving vehicles such as cars, buses, motorcycles, bicycles, trucks, or pedestrians. The death rate due to accidents is three times higher in low-income countries than in high-income countries. The injuries caused by accidents are one of major economic burden in lower economic countries. It is estimated that lower and middle income countries will experience approximately $834 billion dollars (in 2010 USD) in economic losses from 2015–2030 due to fatal and nonfatal crash injuries [2]. The death rate will be reduced if immediate medical care will be provided on time. This would possible if there is any emergency room nearby. The Government should establish emergency rooms around accident zones to reduce death rate due to road accidents. Thus predicting the ideal location for establishing the

 https://doi.org/10.37896/pd91.4/91446

ISSN: 0369-8963

emergency room is the main stay of the research. Some of factors to be considered while choosing a location to set up an emergency room were the proximity to hospitals, dense of accidents, locality, commute time, accessibility and visibility [3]. This project mainly focus on proximity to hospitals and dense of accidents.

Hadi Youzi et al. did a case study optimize on selecting optimal location for hospital in. Kohdasht city using GIS integrated approach and hierarchical analytic process in 2015 based on the criteria such as land prices, distance from other hospitals, access principle, compatible neighborhood that is distance from industrial places, population density, distance from urban center, network access and communication requirements[7]. Christian Stummer et al., integrates hospital plan with multi-objective decision support on determining location and size of medical departments. The proposed two-phase solution using multi-objective Tabu search in the first phase and applying clustering in the second phase and explore the solution space until the “best” configuration is determined [8].

Muhammet Gul et al., presented a literature review on identify the most suitable hospital location to establish a new hospital from the multi-criteria decision-making (MCDM) perspective so that entrepreneurs or government to gain advantage. The preferred reporting items for systematic review and meta-analysis statement (PRISMA) are used as a reference framework. Based on the review, some of the conclusions arrived were Analytic Hierarchy Process **(**AHP) and GIS-based MCDM models plays significant roles in solving this problem, cost, demand, environment, population, government, competition in the market, and distance to some important places are considered the common selection criteria for location selection. In addition to the MCDM structure, the fuzzy structure also used for this purpose [9]. Fatemeh Rahimi et al., in his proposed work conducted studies to assess the location of the existing hospitals in Shiraz using the geographical information system (GIS) and to select optimal locations for establishing new hospitals in Shiraz [10]. Yishu Zhu et al.(2016) implemented a hierarchical location-allocation model, which analyzes the spatial distribution of trauma centers and the candidate locations for trauma centers. Ant colony optimization technique was used to identify the optimal location. Based on the description of a trauma emergency system, optimally locating trauma centers can be considered a hierarchical location problem which includes optimizing the locations of low-level and high-level trauma centers [11].

Yufan Deng et al(2021) developed a model to support location planning in the context of Emergency Medical Service(EMS) in Chengdu, one of the largest cities in southwest China with the population of 16.5 million. The objective his work is aiming to optimize the EMS centers by adding (upgrading) a minimum number of EMS facilities to cover a maximum population. The nearest-neighbor algorithm was used to find the shortest travel time based on geographical information system (GIS) and genetic algorithm (GA) was applied to determine the optimized locations [12].

 https://doi.org/10.37896/pd91.4/91446

ISSN: 0369-8963

Proposed System Methodologies:

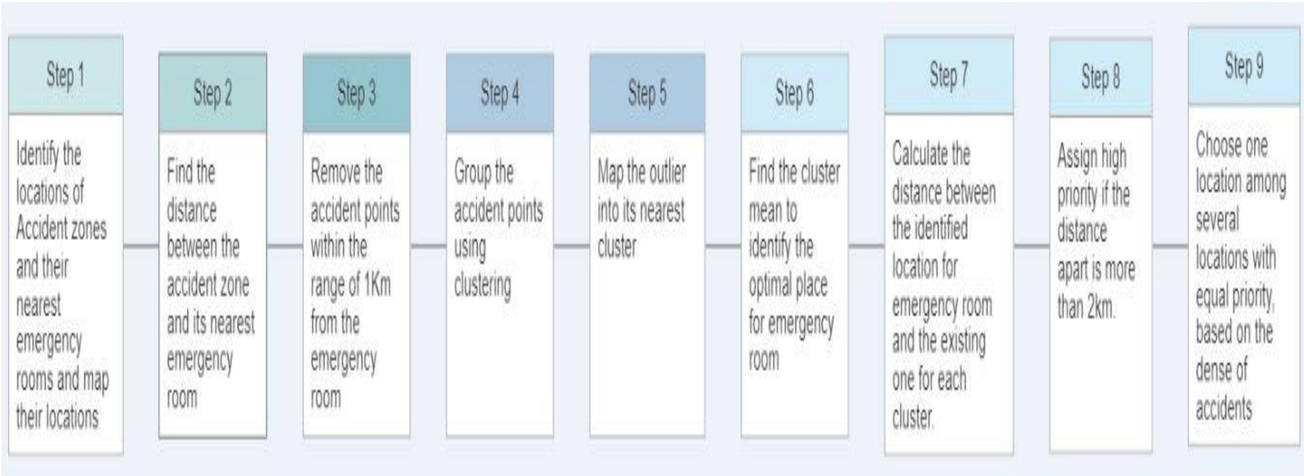

Figure 2.1Block diagram of the Architecture

Algorithm:

Step 1: Identify the locations of accidents zones and the locations of emergency rooms from datasets collected from different sources and map the locations of both.

Step 2: Find the distance between each of the accident zone and its nearest emergency room.

Step 3: Remove the accident points whose distance is within 1Km from the emergency room, as they were already nearby the emergency room.

Step 4: Using HDBSCAN(“Hierarchical Density-Based Spatial Clustering of Applications with Noise*”)* algorithm form the cluster of accident points. The accident points which were farther from the cluster considered as outliers and they are not considered for clustering. Hence considering them may deviate the count of actual number of clusters.

Step 5: The outliers were now mapped into any one of clusters using K-Nearest neighbor algorithm.

Step 6: The optimal place for emergency room to be established for each cluster was identified by taking the mean of the accident points in each of them.

Step 7: The distance between the identified location for emergency room and the existing one was calculated for each cluster to decide the priority among them. The larger the distance between them, the higher the priority. The identified locations which were more than 2Km apart from the existing ones were considered as higher priority and chosen for further processing.

Step 8: Among the several locations with equal priority, only one or two locations was selected based on the dense of accidents occurred in that zone. The total number of emergency room to be established was depends on the fund allocated.

 https://doi.org/10.37896/pd91.4/91446

Methodology:

Data Acquisition:

This work was done by make use two datasets. One was the accident locations of Toronto and the other was locations of emergency rooms in Toronto. Locations of collisions or accidents that happened in Toronto were taken from Toronto public service data portal. It includes the attributes like Most accident stats, date and time, visibility, traffic control, neighborhood district, and locations in terms of latitude, longitude of the accidents, etc. Locations of Hospitals and Emergency rooms in Toronto retrieved using Foursquare API.

Data cleaning:

In the accident dataset the attributes of our interest for this work were the latitude and longitude, and neighborhood district of the accident. From those attributes the density of accidents in that zone was found. From the hospital dataset the latitude and longitude of the hospital, venue name, and type of the venue(medical centre, emergency room, etc), venue id, were taken.

Mapping the two datasets: By mapping the locations of accident zones and the locations of emergency room, the view collision map was found.

Removing the accident points nearby emergency room: The accidents near the hospitals at a distance of 1km are removed from the dataset as they can reach the hospitals in time. The distance of the locations is calculated using the haversine formula. Haversine formula calculates the distance between two points on a sphere when the points are represented using longitudes and latitudes. Hence this work uses geo locations(longitude and latitude), Haversine formula suits well.

$$d = 2r \arcsin\left(\sqrt{\sin^2\left(\frac{\phi_2 - \phi_1}{2}\right) + \cos(\phi_1)\cos(\phi_2)\sin^2\left(\frac{\lambda_2 - \lambda_1}{2}\right)}\right)$$

Identify accident dense area using HDBSCAN: To find the density of the accident zones, the locations of the accidents are grouped together. Grouping is done using HDBSCAN. Advantages of this clustering compared to others are i. more suitable for arbitrary shaped clusters, robust to outliers, does not required to specify the number of clusters.

Three hyper parameters considered to arrive at the optimal number of clusters are epsilon, minPoints / cluster size (n) and minimum samples, where Epsilon (ε), a distance measure that will be used to locate the points in the neighborhood of any point and minPoints(n) is the minimum number of points clustered together to form a region which is considered to be dense and minimum samples, a measure of how conservative the clustering to be. Larger the minimum samples, higher the outliers.

DBSCAN clustering algorithm results in three categories the data point a core point, border point and noise (outlier) point respectively. The point is considered as core, if it has at least n points within distance ε from itself and the data point is considered as border, if it is reachable from Core point

 https://doi.org/10.37896/pd91.4/91446

ISSN: 0369-8963

within the distance ε and the neighborhood contains less than n data points within distance ε, and considered as noise, if it has no core points with the distance of ε[6][7].

HDBSCAN Algorithm:

- The algorithm proceeds by arbitrarily picking up a point in the dataset and labels them into core, border or noise (outlier) and repeat the procedure until all points have been visited.
- Connect the core points which are in neighborhood and put them in a same cluster. Thus a cluster formed with at least one core point, all reachable core points and all reachable border points. If there are at least n points within a distance of 'ε' to the point then we consider all these points to be part of the same cluster.
- The clusters are then expanded by recursively repeating the neighborhood calculation for each neighboring point. Thus, It shapes all the clusters and outliers as well.

The hierarchical dbscan returns label for each data, which represents the cluster in which the data belongs. The label was -1 it the data point was an outlier.

Addressing outliers: The outliers are not considered during HDBSCAN algorithm. So they are not the part of any clusters. k-nearest neighbor's algorithm is used to make the outliers to become the part of the nearest clusters.

Finding the location of emergency room using core latitude and longitude: To find an appropriate location for the emergency room, find the mean for each cluster. This is considered as the core latitude and longitude of the cluster. Using the core latitude and longitude the address of that location was found. This address can be used as an ideal location for setting up an emergency room.

Calculating distance to the nearest medical centre: The distance of the core location and the nearest hospital was found. If the distance between them was more than 2Km, only those points were considered for further processing. The longer the distance between them, the higher the priority while choosing that location. If several locations had equal priority, the number of accidents in that zone was considered as additional parameter on deciding the priority. The number of emergency rooms to be established depends on the fund allocated.

Results and Discussion

The data about collision or accident locations in Toronto were taken from Toronto public service data portal. It includes the attributes like Most accident stats, address of the location such as street with ward number and division number, neighborhood district date and time, visibility, traffic control, , road condition, accident type, impact type, injury and locations in terms of latitude, longitude of the accidents and so on. Data cleaning removes the column with null values and the attributes that are not of much interest. After data cleaning, the table consists of impact type, neighborhood, district, latitude and longitude of the accident location. Locations of Hospitals and Emergency rooms in Toronto retrieved using Foursquare API. It sends back a JSON file from which the dataframe was made. Both the locations, the locations where the collisions occurred and the locations of emergency rooms in Toronto were mapped in Toronto location map. HDBSCAN clustering was done with a minimal sample of 7. If we increase the sample size the cluster becomes

 https://doi.org/10.37896/pd91.4/91446

ISSN: 0369-8963

conservative and results in more number of outliers. The minimal number points to form a dense based cluster were taken as 5. Outliers were mapped into any one of the neighborhood cluster.

	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X
1	DATE	TIME	HOUR	STREET1	STREET2	OFFSET	ROAD_CLASS	District	WardNum	Division	LATITUDE	LONGITUDE	LOCCOORD	ACCLOC	TRAFFCTL	VISIBILITY	LIGHT	RDSFCONI	ACCLA
2	2006/01/0	2325	23	STEELES AV	NINTH LINE ST		Minor Arterial	Scarborou	25	42	43.842745	-79.22479	Mid-Block		No Control	Clear	Dark	Wet	Fatal
3	2006/01/0	1828	18	ISLINGTON	GOLFDOWN DR		Major Arterial	Etobicoke	1	23	43.721445	-79.55809	Intersection	At Interse	Pedestrian C	Clear	Dark	Dry	Fatal
4	2006/01/0	1435	14	KENNEDY F	GLAMORGAN AVE		Major Arterial	Scarborou	21	41	43.769445	-79.28229	Intersection	At Interse	Traffic Signal	Clear	Daylight	Dry	Fatal
5	2006/01/1	1120	11	BARTLEY DI	JINNAH CRT		Collector	North Yorl	16	54	43.722045	-79.30799	Intersection		No Control	Clear	Daylight	Wet	Fatal
6	2006/01/2	1829	18	MIDLAND /	GOODLAND GT		Major Arterial	Scarborough		42	43.779345	-79.27559	Intersection	At Interse	No Control	Clear	Dusk	Dry	Fatal
7	2006/01/2	145	1	WARDEN A	LAWRENCE AVE E		Major Arterial	Scarborou	21	41	43.745345	-79.29469	Intersection		Traffic Signal	Clear	Dawn	Dry	Fatal
8	2006/01/2	754	7	DEAN PARI	BRAYMORE BLVD		Collector	Scarborou	25	42	43.803445	-79.17069	Intersection	At Interse	Stop Sign	Clear	Daylight	Wet	Fatal
9	2006/01/2	2220	22	MOUNT PL	WHITEHALL RD		Major Arterial	Toronto ai	11	53	43.684045	-79.38349	Intersection	At Interse	No Control	Rain	Dark	Wet	Fatal
10	2006/01/2	2048	20	F G GARDII	KIPLING AVE			Etobicoke	3	22	43.618245	-79.52439	Mid-Block		No Control	Rain	Dark	Wet	Fatal
11	2006/02/1	935	9	WESTWOO	WOODYCREST AVE		Local	Toronto ai	14	54	43.687345	-79.34469	Intersection	At Interse	Stop Sign	Clear	Daylight	Dry	Fatal
12	2006/02/2	1840	18	NEILSON R	MCLEVIN AVE		Minor Arterial	Scarborough		42	43.808145	-79.21999	Mid-Block		No Control	Clear	Dusk	Dry	Fatal
13	2006/02/2	501	5	MARTIN GI	RACINE RD		Major Arterial	Etobicoke	1	23	43.717145	-79.58259	Mid-Block		No Control	Clear	Dark	Dry	Fatal
14	2006/03/0	2222	22	F G GARDII	PARK LAWN RD			Etobicoke	3	22	43.627745	-79.48149	Mid-Block		No Control	Clear	Dark	Dry	Fatal
15	2006/03/0	2018	20	F G GARDII	KIPLING AVE			Etobicoke	3	22	43.618845	-79.52039	Mid-Block		No Control	Clear	Dark	Dry	Fatal
16	2006/03/1	852	8	BLOOR ST \	DUNDAS ST W		Major Arterial	Toronto ai	4	11	43.656345	-79.45249	Intersection	At Interse	Traffic Signal	Clear	Daylight	Dry	Fatal
17	2006/03/1	915	9	MORNINGS	SHEPPARD AVE E		Major Arterial	Scarborou	25	42	43.801943	-79.199786	Intersection	At Interse	Traffic Signal	Clear	Daylight	Dry	Fatal
18	2006/03/1	240	2	EGLINTON	COMMONWEALTH A		Major Arterial	Scarborough		41	43.734945	-79.25619	Mid-Block		No Control	Clear	Dark	Dry	Fatal
19	2006/03/1	1535	15	ISLINGTON	DIXON RD		Major Arterial	Etobicoke York		23	43.697045	-79.54669	Intersection	At Interse	Traffic Signal	Clear	Daylight	Dry	Fatal
20	2006/04/1	2000	20	YONGE ST	BREADALBANE ST		Major Arterial	Toronto ai	13	51	43.663745	-79.38409	Intersection		No Control	Clear	Dark	Dry	Fatal
21	2006/04/1	1930	19	WARDEN A	COMSTOCK RD		Major Arterial	Scarborou	20	41	43.721545	-79.28499	Intersection	At Interse	Traffic Signal	Clear	Daylight	Dry	Fatal
22	2006/04/2	843	8	AVENUE RI	CORTLEIGH BLVD		Major Arterial	North Yorl	8	53	43.713045	-79.41179	Intersection	At Interse	No Control	Clear	Daylight	Dry	Fatal
23	2006/04/2	2053	20	KEELE ST	FINCH AVE W		Major Arterial	Etobicoke York		31	43.763445	-79.49099	Intersection	At Interse	Traffic Signal	Clear	Dark	Dry	Fatal

Fatal_Collisions

Figure3.1 Locations of accidents/collisions data set from Toronto public service data portal

Out[5]:

	IMPACTYPE	Neighbourhood	District	LATITUDE	LONGITUDE
0	Approaching	Rouge (131)	Scarborough	43.842745	-79.22479
1	Pedestrian Collisions	Elms-Old Rexdale (5)	Etobicoke York	43.721445	-79.55809
2	Pedestrian Collisions	Dorset Park (126)	Scarborough	43.769445	-79.28229
3	SMV Other	Victoria Village (43)	North York	43.722045	-79.30799
4	Pedestrian Collisions	Agincourt South-Malvern West (128)	Scarborough	43.779345	-79.27559

Figure 3.2 Collision dataset after data cleaning

Figure 3.3 Locations of collisions mapped in Toronto location map

https://doi.org/10.37896/pd91.4/91446


```
Out[170]: [{'reasons': {'count': 0,
    'items': [{'summary': 'This spot is popular',
      'type': 'general',
      'reasonName': 'globalInteractionReason'}]},
   'venue': {'id': '4af2fb96f964a52086e921e3',
    'name': 'Toronto Western Hospital',
    'location': {'address': '399 Bathurst St.',
     'crossStreet': 'at Dundas St. West',
     'lat': 43.65343431584569,
     'lng': -79.40607359183444,
     'labeledLatLngs': [{'label': 'display',
       'lat': 43.65343431584569,
       'lng': -79.40607359183444}],
     'postalCode': 'M5T 2S7',
     'cc': 'CA',
     'city': 'Toronto',
     'state': 'ON',
     'country': 'Canada',
     'formattedAddress': ['399 Bathurst St. (at Dundas St. West)',
```


Figure 3.4 Hospitals location data taken from Foursquare API

Out[15]:

	Venueid	Venue	Latitude	Longitude	Category
0	4af2fb96f964a52086e921e3	Toronto Western Hospital	43.653434	-79.406074	Medical Center
1	4b7d80d9f964a520d3c22fe3	Roswell Park	42.898945	-78.864902	Medical Center
2	4e023cdb6365ba98ee33dd32	Southlake Regional Health Centre	44.061136	-79.452311	Medical Center
3	4b66f884f964a520e6322be3	Buffalo Medical Group (295 Essjay)	42.974214	-78.736627	Medical Center
4	4bd057b177b29c7493848a82	Halton Family Health Centre	43.391872	-79.821928	Medical Center

Figure 3.5 Hospitals location data as dataframe after preprocessing

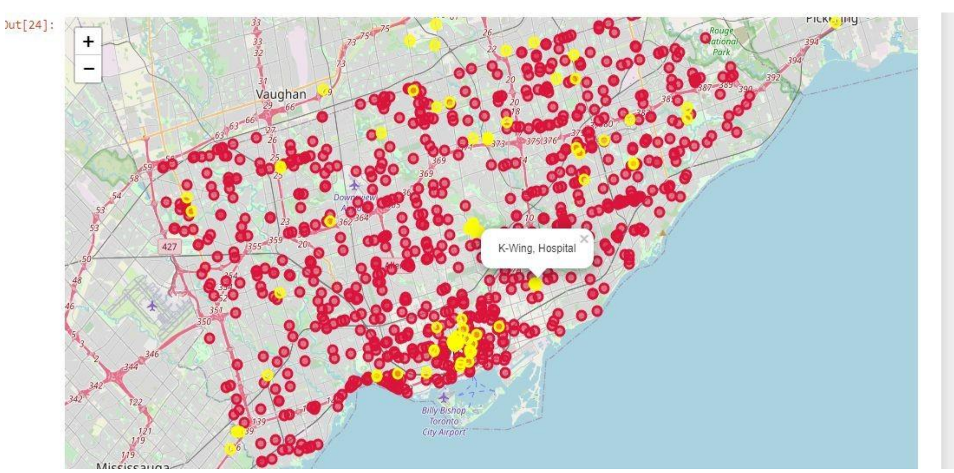

Figure 3.6 Mapping hospitals locations along with collisions location in Toronto map

 https://doi.org/10.37896/pd91.4/91446

ISSN: 0369-8963

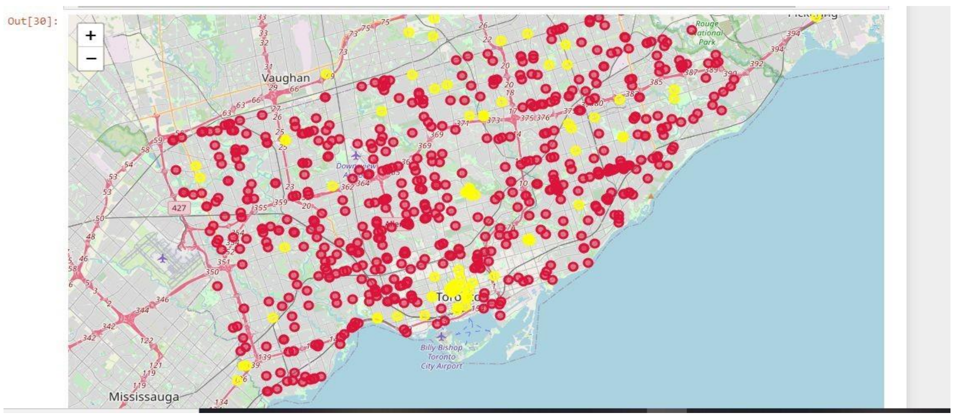

Figure 3.7 Mapping after removing all collisions which are in the radius of 1 Km from any hospital

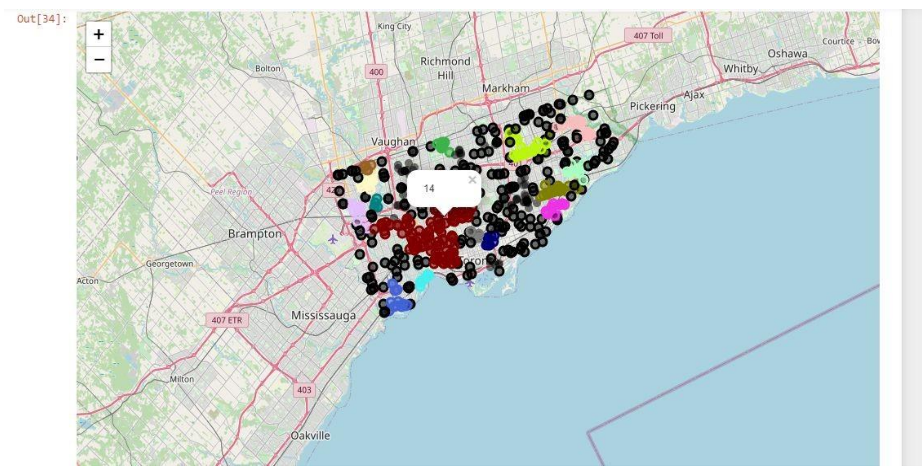

Figure 3.8 Map after clustering collision points based on density using hierarchical DBSCAN

https://doi.org/10.37896/pd91.4/91446

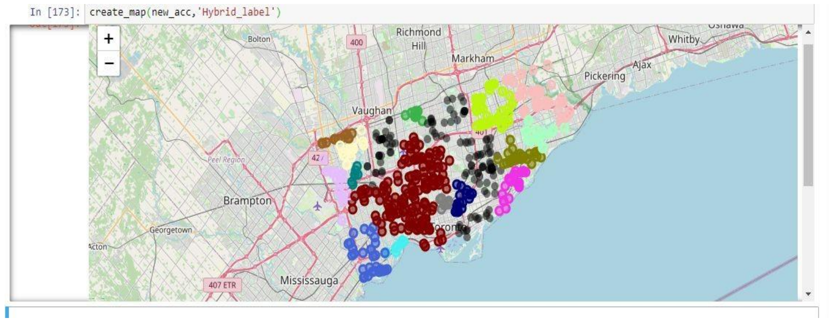

Figure 3.9 Map after clustering the outliers using K-Nearest Neighbors

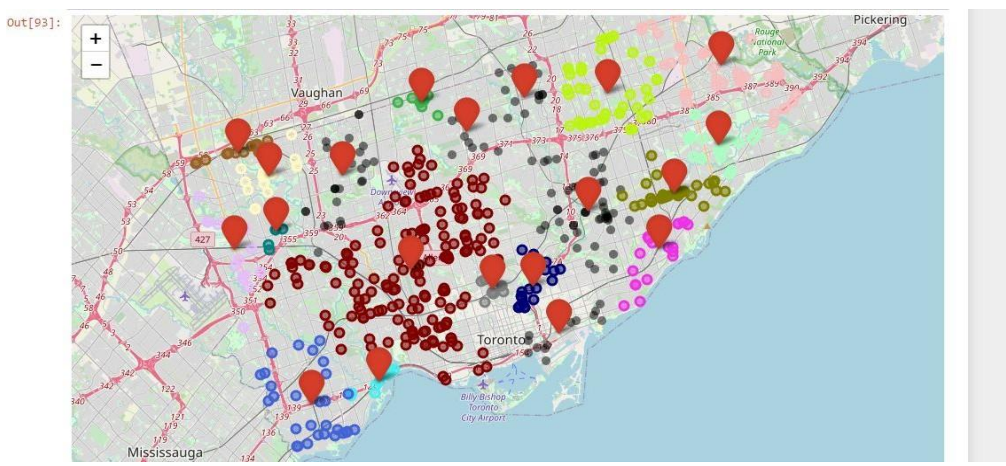

Figure 3.10 Cores points along with their clusters members

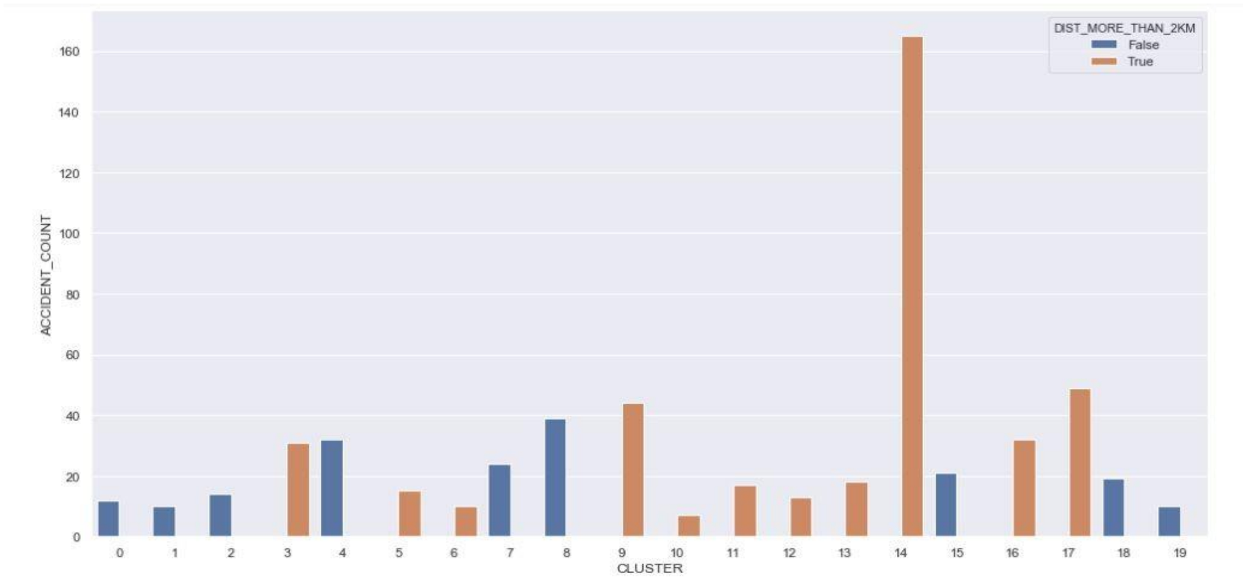

Figure 3.11 Bar plot showing number of accidents in each clusters and distance from each cluster center to the nearest hospital

By considering the number of accidents in each cluster and also the distance to the hospitals, it can be suggested that the cluster with the longest distance and largest number of accidents can be an ideal one for establishing the emergency room.

From the above Bar visualization if we have to select one location then we can choose cluster 14, as it has large number of collisions and also the distance to the nearest hospital from the cluster center is more than 2 Kilometers.

Conclusions

Due to the lack of emergency rooms in the outskirts of the city, and small number of emergency rooms in the central part of the city with high population density, where there are several accident zones. So it is suggested that, in order to access medical facility during emergency in case of accidents especially in areas of lacking healthcare centers with high population density, to establish emergency rooms. In this paper, we have found the optimal place to set up emergency room by considering the number of accident zones and the number of surrounding hospitals and their distances from the accident zones using hierarchical clustering dbscan algorithm which is more suitable for clustering the accident points.

 https://doi.org/10.37896/pd91.4/91446

ISSN: 0369-8963

References

1. https://toronto.citynews.ca/2015/11/26/car-accidents-on-the-rise-in-canada-toronto-study/.
2. https://www.cdc.gov/injury/features/global-road-safety/index.html
3. https://doctors.practo.com/complete-guide-choosing-clinic-location/
4. https://www.mygreatlearning.com/blog/dbscan-algorithm/
5. https://www.kdnuggets.com/2020/04/dbscan-clustering-algorithm-machine-learning.html
6. http://129.122.16.11/bitstream/123456789/3906/1/KWARTENG%20SARAH%20THESIS%202011.pdf
7. Youzi, Hadi & Nemati, Gelareh & Emamgholi, Sara. (2018). The Optimized Location of Hospital Using an Integrated Approach GIS and Analytic Hierarchy Process: A Case Study of Kohdasht City. International Journal of Economics & Management Sciences. 07(01), DOI: 10.4172/2162-6359.1000500.
8. Stummer, C., Doerner, K., Focke, A. et al. (2004), Determining Location and Size of Medical Departments in a Hospital Network: A Multi objective Decision Support Approach. Health Care Management Science 7, 63–71. DOI: https://doi.org/10.1023/B:HCMS.0000005399.23600.69
9. Muhammet Gul, Ali Fuat Guneri, (2021), Hospital Location Selection: A Systematic Literature Review on Methodologies and Applications, *Mathematical Problems in Engineering*, vol. 2021, DOI: https://doi.org/10.1155/2021/6682958.
10. Rahimi F, Goli A, Rezaee R. Hospital location-allocation in Shiraz using Geographical Information System (GIS), Shiraz E-Med J. 2017 ; 18(8):e57572. doi: 10.5812/semj.57572.
11. Yishu Zhu, Qingyun Du et L., Location Optimization Using a Hierarchical Location-Allocation Model for Trauma Centers in Shenzhen, China, International Journal of Geo-Information, 5(10):190, 2016, DOI: 10.3390/ijgi5100190
12. Yufan Deng, Yumeng Zhang, Jay Pan, Optimization for Locating Emergency Medical Service Facilities: A Case Study for Health Planning from China, 2021.