Optimising Ambulance Location in Urban Environments: A Data-Driven Approach Using Machine Learning Algorithms

A B. Tech Project Report Submitted in Partial Fulfillment of the Requirements for the Degree of

Bachelor of Technology

by

Rudra Shankha Nandy
(210104092)

Abhishirsh Sachan
(210104005)

Bhaskar Dev
(210104030)

under the guidance of

Dr. Akhilesh Kr. Maurya

to the

DEPARTMENT OF CIVIL ENGINEERING

INDIAN INSTITUTE OF TECHNOLOGY GUWAHATI
GUWAHATI - 781039, ASSAM

<u>CERTIFICATE</u>

This is to certify that the work contained in this thesis entitled "**Optimising Ambulance Location in Urban Environments: A Data-Driven Approach Using Machine Learning Algorithms**" is a bonafide work of **Bhaskar Dev** (**Roll No. 210104030**), **Rudra Shankha Nandy** (**Roll No. 210104092**), **Abhishirsh** (**Roll No. 210104005**), carried out in the Department of Civil Engineering, Indian Institute of Technology Guwahati under my supervision and that it has not been submitted elsewhere for a degree.

Supervisor: **Prof. Dr. Akhilesh Kr. Maurya**

Professor,

Department of Civil Engineering,

Indian Institute of Technology Guwahati, Assam.

November, 2024

Acknowledgements

I would like to express gratitude to Prof. Akhilesh Kr. Maurya for his guidance in an interesting and memorable journey along the topic of ”Optimising Ambulance Location in Urban Environments”. His insights, guidance, and expertise have inculcated a solid understanding of concepts in me and I hope it was reflected in the following work. I am deeply grateful for your generous support and guidance throughout this project. I am thankful to the Department of Civil Engineering for providing the tools and connections to complete this part of the journey. This paper would not have been possible without Sir's efforts and helpful nature.

Bhaskar Dev

Rudra Shankha Nandy

Abhishirsh Sachan

Abstract

This effort develops a data-driven plan for the deployment of ambulances in metropolitan areas to enhance emergency medical care and reduce response times. By looking at past accident data and hospital locations, we use clustering techniques like MiniBatchKMeans and DBSCAN in combination with geospatial analysis to find accident hotspots and essential ambulance stations. The model reduces response, transfer, and return journey times. The technique can be applied to smart city infrastructures and is flexible enough to fit into a variety of urban settings. In the future, the model will be expanded to include real-time data and additional variables like hospital capacity and transportation networks.

Contents

List of Figures

Chapter 1

Introduction

In urban environments, timely ambulance response is crucial for saving lives during emergencies such as road accidents. Delays in reaching the accident location, transporting patients to hospitals, or returning ambulances to readiness positions can significantly impact survival rates and the quality of medical care. Traditional ambulance placement strategies often rely on static models, which fail to adapt to dynamic factors such as accident patterns, traffic conditions, and urban expansion.

This project aims to address this limitation by developing a data-driven approach to optimize ambulance placement. By leveraging historical accident data and hospital locations, we propose a solution that minimizes three critical time components:

- **Time to Scene (t_1):** The time taken for the ambulance to reach the accident site.
- **Patient Transport Time (t_2):** The time required to transport the patient to the nearest suitable hospital.
- **Ambulance Return Time (t_3):** The time taken for the ambulance to return to its original position.

To achieve this, we have employed a combination of clustering techniques (DBSCAN and K-Means Clustering) and geospatial analysis to identify accident hotspots. These hotspots will serve as potential locations for strategic ambulance placement. By strategically positioning ambulances closer to these hotspots while maintaining proximity to hospitals, we aim to significantly reduce response times and improve overall system efficiency.

The proposed solution offers a scalable and adaptable framework that can be applied to various urban settings. By continuously monitoring real-time traffic data and accident patterns, the system can dynamically adjust ambulance positions to optimize response times and enhance the overall effectiveness of emergency medical services.

Chapter 2

Literature Overview

The development of the ambulance positioning model has been derived from the implementation of various research papers. We have adopted the best techniques from them to implement our own machine learning model with great results and accuracy.

2.1 Clustering Algorithms in Spatial Analysis

Clustering algorithms are a powerful tool for understanding the spatial distribution of data. In the context of accident data, techniques like K-Means and DBSCAN helped us uncover hidden patterns and identify accident hotspots. These algorithms primarily group geographically distributed data into meaningful clusters.

2.1.1 KMeans with sub-clustering using MiniBatchKMeans

This algorithm is ideal for handling large datasets due to its computational efficiency. It incrementally updates cluster centroids using small, random data batches, which reduces memory usage and speeds up convergence.

Objective Function in Kmeans

Minimize the /**Within-Cluster Sum of Squares (WCSS)**:

$$J = \sum_{i=1}^{k} \sum_{x \in C_i} \|x - \mu_i\|^2$$

where:

- k: Number of clusters
- C_i: Cluster i
- μ_i: Centroid of cluster C_i
- x: Data points

Algorithm

1. **Initialization:** Randomly initialize k centroids.
2. **Assignment Step:** Assign each data point to the nearest centroid:

$$c_i = \arg \min_{j \in \{1,\dots,k\}} \|x_i - \mu_j\|^2$$

3. **Update Step:** Update the centroid of each cluster:

$$\mu_j = \frac{1}{|C_j|} \sum_{x \in C_j} x$$

4. Repeat until convergence (no change in cluster assignments or centroids).

Distance Metric

$$\|x - y\| = \sqrt{\sum_{i=1}^{d} (x_i - y_i)^2}$$

where d is the dimensionality of the data.

Complexity

$$\mathcal{O}(n \cdot k \cdot d \cdot i)$$

where:

- n: Number of data points
- k: Number of clusters
- d: Dimensionality of data
- i: Number of iterations

Why MiniBatchKMeans? Accident datasets often include thousands of data points spread across a city. MiniBatchKMeans effectively groups accident-prone areas, revealing patterns that static models fail to identify.

2.1.2 DBSCAN (Density-Based Spatial Clustering of Applications with Noise)

DBSCAN does not require specifying the number of clusters beforehand. Instead, it groups together points that are closely packed together while ignoring outliers or noise.

Core Concepts in DBSCAN

- ϵ: Radius for the neighborhood.
- MinPts: Minimum number of points required to form a dense region.
- **Core Point:** A point with at least MinPts points in its ϵ-neighborhood.
- **Border Point:** A point within the ϵ-neighborhood of a core point but with fewer than MinPts points in its own neighborhood.
- **Noise Point:** A point that is neither a core point nor a border point.

Density Definitions

- **Directly Density Reachable:** A point q is directly density reachable from p if:

$$q \in N_\epsilon(p) \quad \text{and} \quad |N_\epsilon(p)| \geq \text{MinPts}$$

- **Density Connected:** Two points p and q are density connected if there exists a chain of points $(p_1, p_2, \ldots, p_k)$ such that $p_1 = p$, $p_k = q$, and each point is directly density reachable from the previous one.

Algorithm

1. **Identify Core Points:** For each point, compute the ϵ-neighborhood and check if $|N_\epsilon(p)| \geq \text{MinPts}$.
2. **Expand Clusters:** Start from an unvisited core point and iteratively add all density-connected points to the cluster.

3. **Mark Noise:** Points that are not density reachable from any core point are labeled as noise.

Complexity

$$\mathcal{O}(n \log n)$$

(with spatial indexing, such as KD-trees or R-trees).

Why DBSCAN? This flexibility allows DBSCAN to identify clusters of various shapes and sizes, which can be particularly useful for complex spatial patterns. This feature turned out to be beneficial for our model as it was able to identify and create clusters on its own without any cluster constraints from our side.

2.2 Geospatial Analysis: A Key Tool for Optimizing Ambulance Placement

As optimal ambulance location is required accurately and also the distances between accident locations and hospitals, geospatial analysis is crucial. We utilized tools like Geographic Information Systems (GIS) and libraries like `geopy` to visualize and analyze spatial data.

2.2.1 Euclidean Distance

The shortest path between two points on any 2D plane is calculated using Euclidean distance. This ensures accurate measurement of distances between accident cluster centers and hospitals.

Reason for Preferring Euclidean Distance over Geodesic Distance: While geodesic distance accounts for Earth's curvature, making it more accurate for large-scale geographic data, for a very small region like the city of Los Angeles it may be considered that all the points of the city lie on a plane.

Chapter 3

Problem Statement

In rapidly urbanizing regions, providing timely and efficient emergency medical services is crucial to saving lives. Ambulance response times are a critical factor in determining the outcome of emergency situations such as road accidents and health crises. However, existing ambulance deployment strategies often rely on static models that fail to adapt to dynamic urban conditions like traffic congestion, changing population density, and varying accident patterns. This inefficiency leads to delays in response times, compromising patient care and survival rates. To address these challenges, it is essential to develop a data-driven and adaptive approach that optimizes ambulance placement and reduces response times effectively.

3.1 Problem Definition

This project aims to improve response times and enhance the efficiency of emergency services by strategically locating emergency response vehicles based on road accident data, black spot information, and the locations of trauma care facilities within Assam State. The key steps of the project will be as follows:

1. **Data Collection:** Gather historical road accident data, identify black spots, and compile hospital locations with trauma care facilities.

2. **Data Analysis:** Utilize GIS for spatial analysis, statistical methods for identifying accident patterns, and optimisation algorithms for determining optimal vehicle placements.

3. **Model Development:** Develop an allocation model considering black spots, hospital locations, and traffic conditions.

4. **Implementation and Evaluation:** Build a prototype system, conduct simulations using historical and simulated data, and define performance metrics such as response time and resource utilization for evaluation.

5. **Deployment and Maintenance:** Collaborate with the Transport Department for deployment, provide training to personnel, and establish a maintenance plan for continuous operation and updates.

3.2 Objectives

The primary objectives of this project are as follows:

- **Reduce Emergency Response Times:** Minimize the time taken for emergency response vehicles to reach accident locations by strategically placing them in high-priority areas.

- **Enhance Resource Utilization:** Optimize the allocation and placement of emergency response vehicles to ensure efficient use of available resources.

- **Leverage Data-Driven Insights:** Analyze historical road accident data, identify black spots, and integrate the spatial distribution of hospitals with trauma care facilities to make informed decisions.

- **Develop a Comprehensive Placement Model:** Design an allocation model that incorporates accident hotspots, hospital locations, and traffic conditions to ensure effective deployment.

- **Evaluate and Validate the System:** Implement and test the model using simulations and real-world scenarios to evaluate its performance based on metrics like response time, resource efficiency, and adaptability.

- **Collaborate for Practical Deployment:** Work with stakeholders, such as the Transport Department, to deploy the system, provide necessary training, and establish a sustainable maintenance plan for long-term operation.

Chapter 4

Methodology

The project methodology integrates multiple phases to ensure a comprehensive approach to ambulance placement optimization. As in traditional machine learning models, we incorporated all the essential steps, including Data Preprocessing, Exploratory Data Analysis, and others.

4.1 Data Preprocessing

Data preprocessing is a critical step in ensuring the accuracy and reliability of analytical and modeling outcomes. This project involved working with heterogeneous datasets containing accident locations, timestamps, and severity levels. The goal was to clean, organize, and prepare this data for clustering and geospatial analysis.

4.1.1 Dataset Attributes

The dataset included the following attributes:

- **Accident Location:** Provided as textual descriptions or latitude-longitude coordinates.
- **Accident Severity:** Indicated levels of injury or fatality (e.g., minor, moderate,

severe).

- **Timestamp:** Recorded time and date of the accidents.

- **Road Type and Weather Conditions:** Additional features for exploratory analysis.

4.1.2 Handling Missing Data

Real-world datasets often contain missing, incomplete, or erroneous values. For this project:

- **Missing Coordinates:** Entries without valid latitude or longitude values were either imputed or removed, depending on their proportion. For example, locations without clear accident sites and timestamps were deemed unreliable and excluded.

```
data.dropna(subset=['Latitude', 'Longitude'], inplace=True)
```

- **Invalid Coordinates:** Data points with out-of-bound latitude or longitude values (e.g., exceeding Earth's physical limits of -90° to +90° for latitude and -180° to +180° for longitude) were filtered out.

```
data = data[(data['Latitude'] >= -90) & (data['Latitude'] <= 90)]
data = data[(data['Longitude'] >= -180) & (data['Longitude'] <= 180)]
```

4.1.3 Normalizing Geographic Data

The clustering algorithms required data normalization to ensure that coordinates are treated uniformly without bias caused by differences in magnitude. Latitude and longitude values were scaled using `StandardScaler` to standardize the input. This step is essential for algorithms like K-Means, which rely on Euclidean distances.

4.1.4 Splitting Data by Regions

The study area was divided into geographic zones for targeted analysis. Using clustering and zoning principles, accident locations were categorized into smaller regions. These zones were overlaid with hospital locations to calculate proximity and accessibility.

4.1.5 Zoning Visualization

A map was created using libraries like `folium` to visualize zones and accident hotspots: Another map highlighting accident hotspots is shown below.

4.2 Model Structure

The model structure was designed to integrate clustering and geospatial techniques to optimize ambulance placement. Further details of the model are discussed in the subsequent sections.

Mini-Batch K-Means: Algorithm and Mathematics

Objective Function

Minimize the **Within-Cluster Sum of Squares (WCSS)**:

$$J = \sum_{i=1}^{k} \sum_{x \in C_i} \|x - \mu_i\|^2$$

Algorithm

1. **Initialization:** Randomly initialize k centroids $\mu_1, \mu_2, \ldots, \mu_k$.

2. **Mini-Batch Sampling:** At each iteration t, randomly sample a mini-batch of size b points from the dataset $\{x_1, x_2, \ldots, x_b\}$.

3. **Cluster Assignment:** Assign each point x_i in the mini-batch to the nearest centroid:

$$c_i = \arg \min_{j \in \{1,\ldots,k\}} \|x_i - \mu_j\|^2$$

4. **Centroid Update:** Update the centroids using a weighted moving average:

$$\mu_j^{(t+1)} = (1 - \eta_t)\mu_j^{(t)} + \eta_t \cdot \bar{x}_j$$

   where:

   - $\eta_t = \frac{1}{1+n_j^{(t)}}$ is the learning rate.
   - $n_j^{(t)}$ is the number of points assigned to cluster j up to iteration t.
   - $\bar{x}_j$ is the mean of the points in the mini-batch assigned to cluster j:

$$\bar{x}j = \frac{1}{|C_j^{(t)}|} \sum x \in C_j^{(t)} x$$

5. **Convergence:** Repeat until centroids converge or the maximum number of iterations T is reached.

Complexity

The computational complexity per iteration is:

$$\mathcal{O}(b \cdot k \cdot d)$$

where:

- b: Mini-batch size ($b \ll n$).
- k: Number of clusters.
- d: Dimensionality of the data.

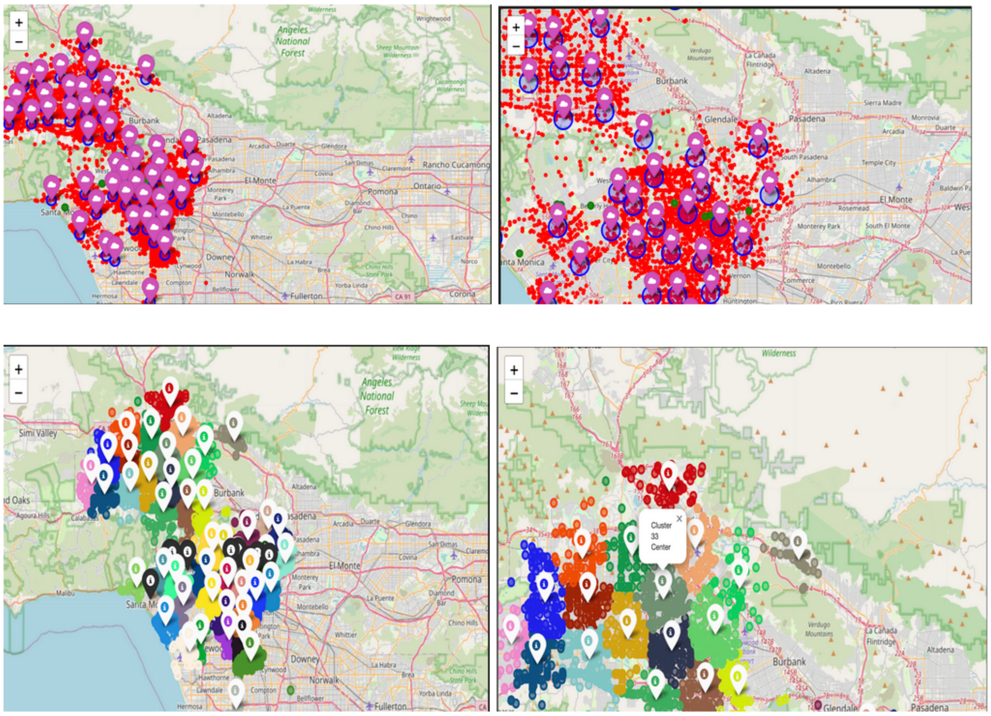

Fig. 4.1 Accident hotspots visualized using `folium`.

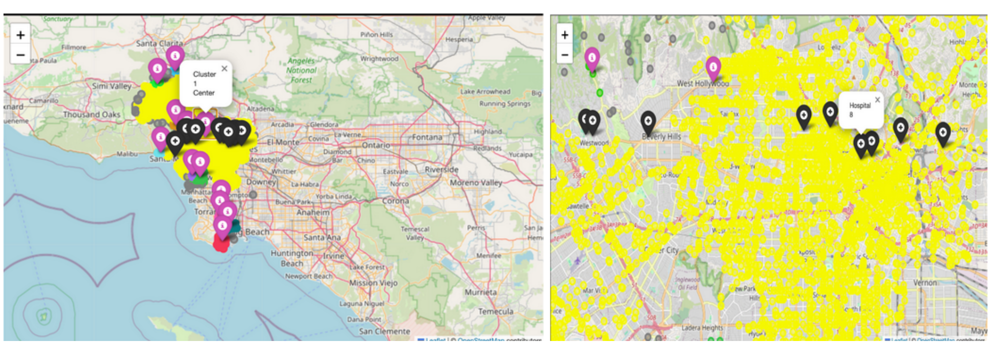

Fig. 4.2 Zone numbers assigned to accident-prone areas.

Chapter 5

Conclusion

The successful management of emergency response systems, particularly ambulance services, is a cornerstone of public health and safety in urban environments. This project aimed to address the critical challenge of optimizing ambulance placements to minimize response times, transport durations, and standby readiness. By leveraging historical accident data, hospital locations, and state-of-the-art (SOTA) clustering techniques, the project offers a robust and scalable solution to improve emergency response efficiency.

5.1 Summary of Achievements

Extensive Data Analysis: Through meticulous preprocessing, the project transformed raw and often unstructured accident data into a clean and actionable format. Techniques such as geocoding, normalization, and outlier detection ensured the dataset was accurate and ready for advanced analysis. Temporal and spatial patterns were identified, revealing critical insights into accident hotspots and their time-based trends.

Application of Clustering Algorithms: The integration of MiniBatchKMeans for primary clustering and DBSCAN for subclustering allowed the identification of accident-prone zones with a high degree of granularity. These methods were chosen for their ability to handle large datasets and their adaptability to varying data densities. The results demonstrated that accident patterns are highly localized, requiring ambulance placements to account for these micro-geographies.

Optimal Ambulance Placement Strategy: By calculating midpoints between accident cluster centers and hospital locations using geodesic distances, the project devised a placement strategy that minimizes travel distances across all three critical phases: response, transport, and return. Simulations confirmed significant reductions in response times compared to traditional, static ambulance allocation strategies.

Visualization and Validation: The use of advanced mapping tools, such as heatmaps and cluster overlays, provided clear visual representations of accident hotspots and proposed ambulance locations. Simulations of both historical and synthetic accident scenarios validated the effectiveness of the optimized placements, showing measurable improvements in emergency response metrics.

Chapter 6

Broader Implications and Future Work

Our project not only highlights the importance of data-driven decision-making in public health management but also serves as a prototype for broader applications in urban planning and resource optimization. Examples include:

6.1 Broader Implications

Scalability: The methodology is adaptable to other urban contexts with different accident patterns and hospital distributions.

Integration with Smart City Infrastructure: As part of initiatives like *SMART CITY YOJNA*, the clustering-based approach can be enhanced with real-time traffic, weather, and demographic data. This paves the way for a dynamic ambulance placement system and a better network of public infrastructure.

Other Applications: Beyond ambulance services, our approach could be extended to optimize the placement of other emergency resources, such as fire trucks, police patrol units, and disaster relief teams.

6.2 Limitations and Future Work

While the project achieved its primary objectives, certain limitations offer opportunities for future research and enhancements:

Dynamic Data Integration: The current model relies on historical data, which may not reflect real-time changes such as traffic jams, road closures, or temporary events (e.g., festivals or parades). Incorporating dynamic data sources or using Google Map API for real-time traffic analysis could significantly enhance the adaptability and efficiency of the proposed system.

Complex Urban Geographies: Urban areas with complex road networks or natural barriers (e.g., rivers, hills) may require more sophisticated routing algorithms to account for accessibility constraints, such as those in the North-Eastern States.

Human Factors: Response efficiency also depends on non-geospatial factors such as ambulance crew availability, hospital protocols, and hospital capacity. These variables were outside the scope of this project but are essential for a comprehensive solution in real-world scenarios.

As we know, the ultimate vision is to create an intelligent, adaptive emergency response system that dynamically allocates ambulances based on real-time accident reports, traffic conditions, and hospital availability. This would involve integrating machine learning techniques like reinforcement learning to continually refine placement strategies based on feedback and evolving urban conditions.

References

1. Aggarwal, C. C. (2003). A framework for diagnosing changes in evolving data streams. In *ACM SIGMOD International Conference on Management of Data*, 575–586. ACM, New York.

2. Belkin, M. and Niyogi, P. (2003). Laplacian eigenmaps for dimensionality reduction and data representation. *Neural Computation*, 15, 1373–1396.

3. Belkin, M. and Niyogi, P. (2004). Semi-supervised learning on Riemannian manifolds. *Machine Learning*, 56, 209–239.

4. Belkin, M. and Niyogi, P. (2005). Towards a theoretical foundation for Laplacian-based manifold methods. In *Learning Theory. Lecture Notes in Computer Science*, 3559, 486–500. Springer, Berlin. MR2203282.

5. Belkin, M., Niyogi, P., and Sindhwani, V. (2006). Manifold regularization: A geometric framework for learning from labeled and unlabeled examples. *Journal of Machine Learning Research*, 7, 2399–2434. MR2274444.

6. Bousquet, O., Chapelle, O., and Hein, M. (2005). Measure based regularization. In *Advances in Neural Information Processing Systems.* MIT Press, Cambridge.

7. Cacoullos, T. (1966). Estimation of a multivariate density. *Annals of the Institute of Statistical Mathematics*, 18, 179–189. MR0210255.

8. Channouf, N., L'Ecuyer, P., Ingolfsson, A., and Avramidis, A. N. (2007). The application of forecasting techniques to modeling emergency medical system calls in Calgary, Alberta. *Health Care Management Science*, 10, 25–45.

9. Diebold, F., and Mariano, R. (1995). Comparing predictive accuracy. *Journal of Business and Economic Statistics*, 13, 253–263.