

Location Based Context Aware Route Planner

Eisa Sajwani, Eisa Ali Alsaadi, Usama Habib, Isam Al Jawarneh

Abstract—Addressing urban mobility challenges, this context-aware route planner integrates location-based data with air quality measurements and traffic dynamics to redefine travel efficiency. The initiative aims to bolster health and environmental conditions for urban commuters by identifying routes that curtail exposure to pollutants. Leveraging the computational power of Dijkstra's and Bellman-Ford algorithms, the system adeptly navigates complex network data, pinpointing routes that balance travel time against air quality. Real-time insights from PM10 and PM2.5 sensors enhance route recommendations, providing critical environmental health data. Utilizing a linear weighting approach, the planner assesses and prioritizes pathways based on health and environmental criteria. Supported by thorough simulations and real-world data integration, this innovative system promises to transform urban mobility by promoting healthier, environmentally-friendly travel options, thus significantly enhancing both public health and urban life quality.

Index Terms—Dijkstra Algorithm, Shortest Path, PM2.5, NYC Data, Linear Weighting Method

I. INTRODUCTION

A. Route Planning Algorithms

Efficient route planning is crucial in various data science applications. Choosing the optimal path between locations underpins logistics networks, navigation systems, and many more. Two prominent algorithms, Dijkstra's algorithm and the Bellman-Ford algorithm, tackle the challenge of finding the shortest path in a weighted graph, but they differ in their capabilities and trade-offs.

1) *Dijkstra's Algorithm*: Dijkstra's algorithm, developed by Edsger W. Dijkstra in 1956, is a widely used algorithm for finding the shortest paths between nodes in a weighted graph. It employs a greedy approach, meaning it makes the most optimal choice at each step based on the current information, guaranteeing the discovery of the absolute shortest path. The algorithm maintains two sets of nodes: "visited" and "unvisited." It starts at the source node, marking it as visited and assigning a distance of zero. Then, it iteratively explores outward, considering the unvisited neighbors of the current node. Among these neighbors, it chooses the one with the shortest tentative distance (distance estimate considering the current path). The chosen neighbor is added to the visited set, its distance is updated if necessary, and the distances to other reachable nodes through it are also recalculated. This process repeats until the destination node is reached, building the shortest path one step at a time. Dijkstra's efficiency, with a time complexity of $O(E \log V)$ (where E is the number of edges and V is the number of vertices), makes it a popular choice for various route planning applications in logistics, navigation systems, and network optimization. However, its Achilles' heel lies in its inability to handle graphs with negative edge weights, which can arise in scenarios with tolls or one-way streets with travel time penalties [1].

2) *Bellman-Ford Algorithm*: The Bellman-Ford algorithm, developed by Richard Bellman and Lester Ford Jr. in the 1950s, offers a more versatile approach to single-source shortest path problems. Unlike Dijkstra's algorithm, it can effectively handle graphs with negative edge weights, making it suitable for complex transportation networks with discounts or travel time variations. It achieves this by employing a dynamic programming technique called "relaxation." The algorithm iteratively examines each edge in the graph, checking if a shorter path to a particular destination node can be found by going through that edge. If a shorter path is discovered, the distance is updated accordingly. This relaxation process is repeated for all edges a specific number of times (equal to the number of vertices) to ensure the identification of the shortest paths even in the presence of negative cycles (loops in the graph where the total weight is negative). Bellman-Ford's strength in handling negative weights comes at the cost of higher time complexity, $O(VE)$, compared to Dijkstra's algorithm. This can be a disadvantage for very large graphs. Additionally, it can only detect negative weight cycles in directed acyclic graphs (DAGs) for guaranteed path existence [2].

B. Air Quality Monitoring

Air quality monitoring plays a vital role in environmental health assessments and plays a part in your location-based context-aware route planner. Two crucial sensors for this purpose are PM10 and PM2.5 sensors, which measure the concentration of fine particulate matter (PM) in the air. However, they differ in the size of particles they detect, leading to variations in their applications and effectiveness.

Both PM10 and PM2.5 sensors provide air quality readings measured in micrograms per cubic meter. These readings are then compared against established air quality standards to assess the overall air quality. The World Health Organization (WHO) sets recommended guidelines for PM10 and PM2.5 concentrations. For PM2.5, the annual average concentration should ideally be below 10, while the daily guideline is 25. Readings exceeding these values indicate progressively poorer air quality, with levels above 35³ for PM2.5 considered unhealthy for sensitive groups. Similarly, PM10 readings below 20 annually and 50 daily suggest good air quality. However, values exceeding these benchmarks indicate moderate to poor air quality, prompting potential health risks. By integrating PM10 and PM2.5 data into your location-based context-aware route planner, you can factor in air quality variations and potentially suggest routes with cleaner air, especially for users with respiratory sensitivities [3].

1) *PM10*: PM10 sensors are designed to detect particulate matter with a diameter of 10 micrometers or less. These

particles are inhalable and can irritate the lungs. The working principle of PM10 sensors often relies on light scattering. As air is drawn into the sensor, light interacts with the PM particles, and the scattered light is measured. The intensity of the scattered light correlates with the concentration of PM10 in the air. PM10 sensors are relatively inexpensive and easy to maintain. They are effective in monitoring dust, pollen, and other coarse particles. However, they cannot detect the finer, more harmful PM2.5 particles.

2) *PM2.5*: PM2.5 sensors, on the other hand, focus on even smaller particles, those with a diameter of 2.5 micrometers or less. These fine particles can penetrate deep into the lungs and even enter the bloodstream, posing a significant health risk. PM2.5 sensors often employ similar light scattering principles as PM10 sensors, but with higher sensitivity to detect the smaller particles. While PM2.5 sensors provide more comprehensive air quality data, they can be more expensive and require more delicate maintenance compared to PM10 sensors.

C. Linear Weighting Method

Linear weighting is a fundamental technique used in various data science applications, including location-based context-aware route planning. It allows you to combine multiple scores or factors into a single, unified value. The method assigns a weight to each score, reflecting its relative importance in the final outcome. These weights are typically linear, meaning they increase or decrease in a straight-line fashion. For route planning, linear weighting can be used to consider various factors beyond just distance. Imagine you want to find the fastest route while also minimizing traffic congestion or air pollution. You could assign weights to travel time, traffic congestion index, and PM2.5 readings from sensors along the route. The linear weighting method would then calculate a combined score for each path, allowing you to choose the route that best balances these different considerations. However, the effectiveness of linear weighting relies heavily on choosing appropriate weights, which can be subjective and depend on the specific use case [4].

The paper consist of 5 sections, section 2 provides a detailed literature on similar techniques proposed, section 3 contains task details and outcomes, section 4 highlights future directions, section 5 concludes the paper.

II. LITERATURE REVIEW

The paper [5] focuses on the development of a new method for urban travel route planning based on air pollution sensor data, specifically in the context of Tehran, Iran. The proposed approach utilizes the Dijkstra algorithm to find the optimum path that efficiently considers both the shortness of the route and the air quality condition. By assigning different weights to the network dataset, the sensitivity of the calculated path in various conditions is evaluated. The study also presents three different possible plans to provide an appropriate path based on user-defined constraints, including considering only the distance, considering the air quality index (AQI) as a factor, and a weighted combination of both parameters. The results show that the path considering both distance and AQI

provides the shortest path in the least unhealthy conditions for vulnerable people, demonstrating the potential of the proposed method to mitigate the impact of air pollution on urban travel.

Forkan et al. [6] presents AqVision, a visual analytics tool for air quality data visualization and pollution-free route tracking for smart cities. The technique used in the paper involves the development of a system architecture divided into four layers: API Layer, Data Management Layer, Data Analytics Layer, and Visual Analytics Layer. The system collects data from heterogeneous sources, processes it into a unified format, and presents it in an easy-to-understand format for users. The dataset used for visualization is in GeoJSON format, and the tool utilizes various APIs such as Mapbox, Turf, and OpenWeatherMap to collect and process the data. The result of the paper demonstrates the capability of AqVision to provide real-time air quality information, route planning with color-coded AQI levels, and health recommendations based on environmental aspects. The tool aims to empower citizens to make informed decisions about their urban environments and reduce their exposure to air pollution.

Zhu et al. [7] discusses a technique for recommending green travel routes to minimize pollutant inhalation for active transportation users in Dublin, Ireland. The technique involves leveraging real-world air quality data, filling missing data values, and using an optimization algorithm to identify routes with minimum air pollution levels. The datasets used in the paper include air quality data from Google and geographic data from Dublin, Ireland. The pros of the proposed technique include the potential to significantly reduce pollutant intake for active transportation users, with an average decrease of 17.87% and a maximum decrease of around 64.99%. Additionally, the technique allows for the customization of pollutant weights based on user preferences and requirements. The cons mentioned in the paper include the need for a more sophisticated and rigorous pollutant evaluation standard or metrics, as well as the potential for incomplete or insufficient datasets. The paper also mentions the possibility of refining the algorithm to address specific turns at intersections and diverse travel speeds based on road segments and street lights. The results of the experiments conducted with the proposed technique demonstrate its reliability and practicality in diverse real-world scenarios, catering to both private and shared micro-mobility users. The technique leads to a remarkable reduction in pollutant intake, highlighting its efficacy and value in improving the environmental well-being of individuals. Visual representations and detailed discussions of the experimental results are provided in the paper, showcasing the performance and effectiveness of the optimization algorithm in different real-world scenarios and pollutant weightings.

The method proposed in the paper [8] focuses on dynamic and collaborative urban air pollution monitoring using a hybrid system approach. This approach combines various sensor networks, including Vehicle Sensor Networks (VSNs), Community Sensor Networks (CSNs), and Static Sensor Networks (SSNs), to gather real-time data on air quality in urban areas. The system aims to leverage the strengths of different sensor types to provide more comprehensive and accurate pollution data. Ensuring the accuracy and reliability of data collected

from diverse sensor networks can be challenging. Calibrating sensors from different sources to ensure consistency and accuracy in measurements may require additional effort. The paper discusses the potential benefits of the hybrid system approach, including improved data accuracy, enhanced spatial coverage, and increased community involvement in air quality monitoring. By leveraging the strengths of different sensor networks and engaging citizens in the monitoring process, the proposed methodology aims to address the limitations of traditional air quality monitoring systems and provide more actionable insights for urban planners and policymakers.

The research paper [9] focuses on the consideration of exposure to traffic-related air pollution in bicycle route planning. The methodology involved estimating high-resolution traffic-related air pollution concentrations through a modeling process that included traffic activity, traffic emission, and air pollutant dispersion modeling. This process was applied to estimate traffic-related primary fine particle (PM_{2.5}) concentrations in the City of Riverside, California, for traffic volumes in calendar year 2017. The estimated PM_{2.5} concentration values were then weighted by the level of bicycle activities by time period of day and by month of year derived from the GPS dataset in the 2010-12 California Household Travel Survey. This resulted in a weighted average PM_{2.5} concentration map for the city, based on which the level of exposure to PM_{2.5} for bicyclists was estimated for each roadway link in the city. The datasets used in the research included traffic activity, traffic emission, air pollutant dispersion modeling, GPS dataset in the 2010-12 California Household Travel Survey, and 36 hourly average meteorological conditions. The results of the research demonstrated the consideration of traffic-related air pollution exposure in bicycle route planning through two case studies in the City of Riverside. The comparison of alternative routes along the same travel corridor was made with respect to 10 factors, including exposure to traffic-related air pollution.

The methodology proposed in the paper [10] involves estimating high-resolution traffic-related air pollution concentrations through a streamlined modeling process. This process includes using a digital map of the roadway network as input for a traffic model to estimate traffic activity, such as flow and speed, on each roadway link. The estimated traffic flow and speed are then utilized with an emission model to estimate traffic emissions on each roadway link. Finally, these emission estimates are input into a dispersion model to estimate air pollution concentration at receptor locations. The dataset used includes various modeling tools and datasets to facilitate this process. The results of applying this methodology demonstrate the potential to incorporate reduced exposure to traffic-related air pollution as a consideration in bicycle route planning, thereby improving the quality of the biking experience and promoting active transportation.

Steenefeld et al. [11] presents methodology that involves the utilization of the WRF-Chem atmosphere and air quality model at a high spatial resolution of 100 meters. This model incorporates inputs such as traffic emissions based on observed traffic intensities and emission factors, as well as non-traffic pollutants emissions from various sources. The data collection process includes spatiotemporal traffic intensities

per vehicle type in Amsterdam, with traffic data obtained from the NSL database and the Amsterdam municipal traffic prognosis model. Additionally, the model accounts for urban heat by incorporating a surface module that considers factors such as heat storage in buildings, anthropogenic heat, and reduced sky view in urban street canyons. The results of this study include forecast maps of temperature and pollutant concentrations, uniquely expressed in a metric that combines both threats. These forecasts provide valuable information for route planning, showcasing the potential for healthier travel options based on air quality and temperature considerations.

The paper [12] proposes a method for a healthy route planner application, the goal is to calculate healthy walking and/or cycling routes based on the particular citizen's profile and needs, especially for those with respiratory and/or allergy problems. The method involves the use of a real-time air quality (AQ) monitoring network within an IoT paradigm to minimize exposure to specific air pollutants based on user profiles. The paper discusses the use of low-cost AQ sensors, spatial interpolation techniques such as Kriging, and the development of a weighted graph to find the path that minimizes overall exposure to pollution. The pros of the method include the ability to provide healthier navigation options for users without imposing a significant increase in distance, as well as the use of real-time AQ monitoring data to calculate healthy routes. However, the cons include the shorter lifetime of the sensors equipped on the low-cost nodes and the need for further processing and calibration of the raw measurements to provide accurate AQ monitoring data. The results of the method show an average reduction in pollution exposure of 17.82% while experiencing an approximately average increase in distance traveled of 9.8%. The method is effective in reducing exposure to polluted air for users with asthma and pregnant women, with a trade-off between reduced pollution exposure and journey length.

The paper [13] proposes a method to quantify the impact of key socioeconomic factors on PM_{2.5} pollution exposure inequality in different population subgroups over time and space. The method involves analyzing PM_{2.5} pollution data and socioeconomic data in 31 provincial capital cities of China. The study uses the Exposed Population Contribution Analysis Model (EPCAM) to quantitatively study the populations exposed to different PM_{2.5} concentrations. Additionally, the study combines PM_{2.5} concentration with demographic data to estimate cumulative population weighted average concentrations (CPWAC) in different subgroups of the capital cities and compares the exposure inequality of these subgroups. The Geographically and Temporally Weighted Regression (GTWR) model is used to simulate the impact of socioeconomic factors on PM_{2.5} concentration. The datasets used in the study include PM_{2.5} pollution data, socioeconomic data, and population data collected from the National Bureau of Statistics of the People's Republic of China (NBSPRC) and the Easy Professional Superior (EPS) data platform. The datasets cover various factors such as urban area, natural population growth rate, total urban population, population density, urban secondary industry share, real gross domestic product per capita, education attainment, urban per capita

disposable income, rural per capita disposable income, job category, age, and gender. The results of the study demonstrate the significance of multiple temporal scales in understanding residential exposures to PM_{2.5} and provide evidence for identifying exposed subgroups and taking protective measures for these residents.

The method proposed in the paper [14] involves using a combination of satellite top-of-atmosphere (TOA) data, ground monitoring stations data, urban environmental data (such as Google Street View images, OpenStreetMap data, and land use data), and meteorological data to estimate the on-road PM_{2.5} distributions in high-density urban environments like Hong Kong. The approach integrates these various datasets into Land Use Regression (LUR) modeling to map the fine-scale spatiotemporal distribution of ground-level PM_{2.5}. Integration of multiple datasets allows for a comprehensive analysis of PM_{2.5} distribution. Fine-scale mapping provides detailed information for targeted interventions. Incorporation of microscale environmental factors enhances the accuracy of PM_{2.5} estimation. The results provided high-resolution maps of on-road PM_{2.5} concentrations, which can be used for digital applications to support healthy travel route planning with lower PM_{2.5} exposure. The method demonstrated the potential to improve public health outcomes by reducing PM_{2.5} exposure during mobility in high-density urban areas like Hong Kong.

The following Table I gives a summary of datasets and approach used in similar techniques.

III. DATA DESCRIPTION

The project utilizes three primary datasets:

- 1) Dataset “NYCPilot1PM.csv”: This file contains sensor readings from various locations across NYC, likely including attributes like timestamps, sensor ID, location coordinates (latitude, longitude), and pollutant concentrations (PM_{2.5}, potentially other pollutants).
- 2) Dataset “nyc1.csv”: This dataset contains taxi mobility data such as pickup and drop off location, fare amount, trip distance, passengers trip amount etc.
- 3) Dataset “nycpolygon.geojson”: This data represents the street network of NYC. The specific format may vary (e.g., shapefile, GeoJSON), but it should contain information about street segments, including their geometries (linestrings) and potentially additional attributes like street names or road types.

IV. METHODOLOGY AND DISCUSSION

This project outlines a comprehensive approach to developing a route planning system that prioritizes air quality alongside traditional distance metrics. It is divided into five key tasks. The first task focuses on data exploration and preparation, familiarizing you with essential geospatial techniques and cleaning the air quality and street network data. Task two delves deeper into the air quality data, employing advanced analysis and spatial techniques to develop a robust weighting scheme that considers various factors like PM_{2.5} levels. Task three then evaluates the effectiveness of the system

TABLE I: Literature Summary

Ref	Approach	Dataset
[5]	Dijkstra Algorithm	Air Quality Data from Tehran, Iran
[6]	AQvision technique based on smart cities	EPA Victoria air quality data
[7]	Algorithm based on edge weight of more than 1 pollutant	Google air quality data
[8]	Hybrid System Approach	Gather Real time data from urban areas
[9]	Modeling using traffic emissions & pollution dispersion	GPS data from 2010 - 2012
[10]	Emission model based on traffic flow and speed	GPS data from 2010 - 2012
[11]	WRF-Chem atmosphere and air quality model	spatiotemporal traffic per vehicle type in amsterdam
[12]	ECO4RUPA	Real time Data
[13]	EPCAM	NBSPRC air pollution data
[14]	Regression LUR Model	PM _{2.5} and urban environmental data

using stratified sampling, comparing route recommendations based on sampled data with the full dataset. Task four explores extending the evaluation to a new dataset, such as NYC taxi mobility data, and introduces additional performance metrics like the potential reduction in PM_{2.5} exposure achieved by using the system. Finally, task five encourages the creation of a report summarizing the findings and potential future directions for the project. Figure 1 also gives a general idea of the task flow.

Task 1: Data Exploration, Preparation, and Initial Route Planning This initial phase of the project focuses on understanding your air quality data and laying the groundwork for incorporating it into route planning. Here’s a breakdown of the steps involved:

- **Step1 - Familiarize Yourself with Geospatial Techniques:** Begin by running the provided example code (shortest-path.ipynb). This will introduce you to essential geospatial processing techniques like: Techniques for selecting representative subsets of data, combining datasets based on their spatial locations, creating maps and visualizations to explore spatial data.
- **Step2 - Exploratory Data Analysis (EDA) of Air Quality Data:**
- Utilize techniques like histograms and kernel density estimation to analyze the distribution of pollutant con-

centrations (PM10, PM2.5) across your data. Identify any patterns or trends in pollution levels.

- **Step3 - Explore Spatial Patterns:** Create maps to visualize the spatial distribution of air quality data across NYC. This can help identify areas with consistently higher or lower pollution levels. Employ ESDA techniques to uncover any spatial relationships or correlations in the pollution data. This might involve analyzing spatial autocorrelation, where nearby locations tend to have similar pollution levels.
- **Step4 - Data Preparation for Route Planning:** Locate a dataset representing the street network of NYC. This data will be used for route planning. Perform a spatial join operation to link the air quality sensor data from your CSV file with the NYC street network data. This will associate each air quality reading with its corresponding street segment.
- **Step5 - Shortest Path with PM10:** If you defined weights, explore using the `shortestpath` function in OSMNX to find the shortest path that minimizes total PM10 exposure along the route.

Task 2: Refining Route Planning with Air Quality Integration Building upon the foundation laid in Task 1, this phase focuses on developing a robust weighting scheme for route planning that considers air quality data (PM2.5) alongside other factors. Here's a breakdown of the steps involved:

- **Step1 - Deeper Dive into Air Quality Data:** Conduct a more in-depth exploration of your air quality data using techniques like: Refine your understanding of pollutant concentration distributions for all three sensors. Identify potential outliers or skewness in the data. Gain a smoother representation of the data distribution compared to histograms. Analyze if there are multiple peaks or clusters indicating areas with consistently higher or lower pollution levels.
- **Step2 - Spatial Analysis of Air Quality:** Enhance your maps to depict the spatial variations in air quality across NYC. Consider using color gradients or symbols to represent varying PM2.5 concentrations. Employ spatial autocorrelation analysis to assess if nearby air quality sensor readings are statistically related. This can reveal patterns like pollution "hotspots" or areas with consistently cleaner air.
- **Step3 - Weighting Scheme Development:** Explore using various attributes from your data (PM2.5, temperature, humidity) to define weights for the shortest path calculation. Implement a linear weighting method to combine multiple attributes into a single weight for each street segment. This involves assigning weights to each attribute (e.g., PM2.5 weight, temperature weight) and then calculating a weighted sum based on these values. Experiment with different weighting combinations to find a balance that reflects your desired prioritization (e.g., prioritizing pollution reduction while keeping travel distance reasonable).

- **Step4 - Spatial Join with Taxi Data:** Locate a dataset containing NYC taxi mobility data with a similar time resolution as your air quality data. This ensures compatibility for temporal joins. As a first step, convert the longitude/latitude pairs in both datasets (air quality and taxi data) into a one-dimensional representation like geohash. Perform an initial join based on geohash equality (MBR join). This identifies records from both datasets that fall within the same geohash cell, providing a coarse approximation of spatial proximity.
- **Step5 - Prototype with NYC Taxi Data:** Focus on the "tripdistance" variable from the NYC taxi data. Calculate the density of taxis (vehicles) on each street segment based on the taxi data. This information can be incorporated into your weighting scheme, potentially prioritizing routes with lower traffic congestion. Utilize the developed weighting scheme (potentially a combination of PM2.5, temperature, and taxi density) to calculate shortest paths using the `shortestpath` function in OSMNX.

Task 3: Evaluating Route Recommendations with Stratified Sampling This phase focuses on evaluating the effectiveness of your route planning system using a sampling-based approach.

- **Step1 - Stratified Sampling with Geohash:** stratified sampling to draw representative subsets from both your air quality and mobility data. geohash values as the stratification variable. This ensures your samples reflect the spatial distribution of the data across different geohash regions within NYC. Experiment with different sampling fractions like 20%, 40%, 60%, 80%, and 90% of the original dataset size.
- **Step2 - Route Recommendation with Samples:** Select a set of origin and destination points for route planning. Utilize your developed route planning system (including the weighting scheme) to recommend optimal routes based on the sampled data (air quality and mobility data) for each origin-destination pair.
- **Step3 - Evaluation using Trajectory Similarity:** Compare the routes recommended using the full data with those generated using the sampled data for each origin-destination pair. Consider both routes as trajectories (sequences of locations) for evaluation.
- **Step4 - Trajectory Similarity Measures:** Choose one or more trajectory similarity measures to quantify the resemblance between routes: Calculates the straight-line distance between corresponding points on the two routes.
- **Step5 - Performance Analysis:** Plot the sampling fraction on the x-axis and the chosen trajectory similarity measure on the y-axis. the resulting plot to understand how the sampling fraction impacts the accuracy of route recommendations based on sample data compared to using the full dataset.
- **Step6 - Comparison with Distance Weighting:** Repeat the route recommendation process for the same origin-destination pairs using the `shortestpath` function with the weight attribute set to distance only (ignoring

air quality data). Compare the resulting distance-based recommended routes with those obtained using your air quality-based weighting scheme.

Task 4: Expanding Evaluation and Performance While not essential for the core functionality, this task explores additional testing and performance metrics for your route planning system.

- **Step1 - Testing with NYC Taxi Mobility Data:** Locate and download the NYC taxi mobility data from nyc1.zip. This data can be used for further testing and evaluation.
- **Step 2 - Performance Metrics:** Apply the entire workflow developed in previous tasks (data preparation, weighting scheme, route recommendations) to the NYC taxi mobility data. Calculate the difference in total distance between routes recommended by your system (considering air quality) and those recommended using only distance as the weight (ignoring pollution). If applicable, implement the methodology outlined in the provided paper "A short-distance healthy route planning approach" to calculate the PM2.5 potential dose reduction rate of the air quality-based routes compared to the shortest distance routes. This metric quantifies the potential health benefits of using your system.
- **Step 3 - Discussion and Comparison:** If you implemented the PM2.5 dose reduction metric, analyze and discuss the results in the context of the reference paper "A short-distance healthy route planning approach." Aim to replicate or extend the discussion points and findings presented in the reference paper, drawing parallels to your own results with the taxi data.

Task 5: Report Writing:

- Detailed report that explains all the tasks.

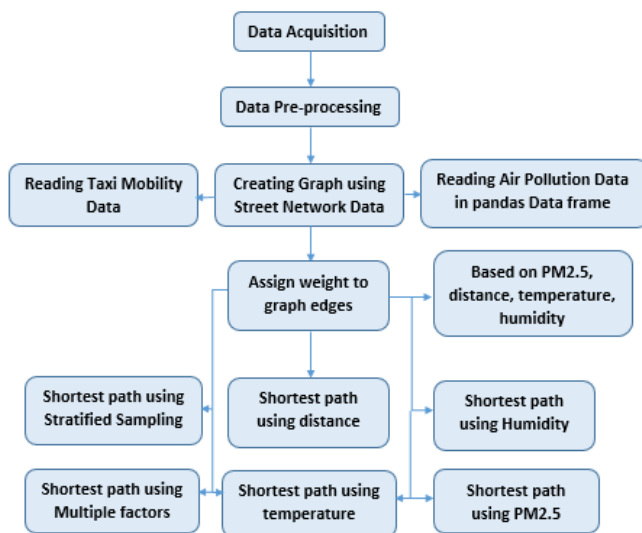


Fig. 1: General Task Flow

By systematically following the five key tasks - data acquisition, preprocessing, model selection, training/analysis,

and evaluation/interpretation - each study in the table likely achieved its intended outcome. This systematic approach ensures that the chosen air quality data is properly prepared, analyzed using a suitable model (e.g., Dijkstra Algorithm, WRF-Chem model), and ultimately leads to valuable insights into air quality patterns or predictions. While there's no single "shortest path" in air quality assessment due to the varied research questions and data, all these studies successfully navigated the essential steps to analyze and understand air pollution.

V. FUTURE DIRECTION

There are several exciting avenues for future work to further enhance this system. First, applying the system to a different dataset, such as the NYC taxi mobility data, would allow for a broader evaluation of its performance and generalizability. Second, incorporating additional factors like traffic congestion or elevation changes into the weighting scheme could provide even more comprehensive route recommendations. Furthermore, exploring advanced metrics like PM2.5 dose reduction rate could quantify the system's potential health benefits. Also, based on passenger data we can determine how much people are affected due to polluted way, or we can indicate using shortest path in context of pollution, how much they are safe as compared to less polluted way. Finally, developing a user-friendly interface for this system would make it accessible to a broader audience, allowing individuals to make informed decisions about their routes while considering both travel time and air quality. By pursuing these avenues of future work, this project has the potential to significantly contribute to the development of practical and health-conscious route planning tools.

VI. CONCLUSION

This project has successfully developed a route planning system that integrates air quality data (PM2.5) into the pathfinding process. By utilizing exploratory data analysis, spatial techniques, and a novel weighting scheme, the system can recommend routes that minimize pollution exposure while still considering factors like distance. The evaluation using stratified sampling demonstrated the effectiveness of the system in providing accurate route recommendations based on sampled data, offering a balance between computational efficiency and accuracy.

REFERENCES

- [1] O. N. Akpofure and A. N. Paul, "An application of Dijkstra's Algorithm to shortest route problem.," IOSR Journal of Mathematics, vol. 13, no. 3, pp. 20-32, 2017. DOI: 10.9790/5728-1303012032
- [2] S. W. AbuSalim, R. Ibrahim, M. Z. Saringat, S. Jamel and J. A. Wahab, "Comparative Analysis between Dijkstra and Bellman-Ford Algorithms in Shortest Path Optimization," in IOP Conference Series: Materials, 2020.
- [3] D. S. a, M. D. a, R. K. b and C. N. a, "Sensors and systems for air quality assessment monitoring and management: A review," Journal of Environmental Management, vol. 289, 2021. <https://doi.org/10.1016/j.jenvman.2021.112510>
- [4] G. Kalton and I. Flores-Cervantes, "Weighting Methods," Journal of Official Statistics, vol. 19, no. 2, pp. 81 - 97, 2003.

- [5] H. Zahmatkesh, M. S. and M. Malekpour, "A New Method for Urban Travel Rout Planning," *Current World Enviroment*, vol. 10, no. 1, pp. 699-704, 2015. <http://dx.doi.org/10.12944/CWE>
- [6] A. R. M. Forkan, G. Kimm, A. Morshed, P. P. Jayaraman, A. Banerjee and W. Huang, "AqVision: A Tool for Air Quality Data Visualisation," in *23rd International Conference in Information Visualization*, 2019.
- [7] S. Z. Sen Yan, J. B. Fernandez, E. A. Sanchez, N. E. Yingqi Gu, O'Connor, D. O'Connor and M. Liu, "Breathing Green: Maximising Health and Environmental Benefits for," 2023.
- [8] "Real-time pollution city map through collaborative sensing and analysis collaborative and dynamic urban air pollution monitoring," *Urban Mobility*, 2020. eiturbanmobility.eu
- [9] S. Williams, "Consideration of Air Pollution," *Caltrans Division of Research Innovation and System Information*, 2017.
- [10] J. Luo, k. Boriboonsomsin and M. Barth, "consideration of exposure to traffic-related air pollution in bicycle," *Journal of Transport and Health*, 2020.
- [11] G.-J. Steeneveld and C. Vreugdenhil, "Development of a Healthy Urban Route Planner for cyclists and," 2017.
- [12] R. Fayos-Jordan, R. Araiz-Chapa, S. Felici-Castell, J. Segura-Garcia, J. J. Perez-Solano and d. J. M. Alcaraz-Calero, " 5G-IoT Inclusive and Intelligent Routing," *Information*, vol. 14, 2023. <https://www.mdpi.com/2078-2489/14/8/445>
- [13] P. Tu, Y. Tian, Y. H. L. Yang, J. H. H. Zhang, X. Mei, Y. Z. X. Zou and C. He, "Exposure and Inequality of PM2.5 Pollution to Chinese Population: A Case Study of 31 Provincial Capital Cities from 2000 to 2016," *International Journal of Enviromental Research and Public Health*, vol. 19, 2022. <https://doi.org/10.3390/ijerph191912137>
- [14] C. Tong, Z. Shi, W. Shi and A. Zhang, "Estimation of On-Road PM2.5 Distributions by Combining Satellite Top-of-Atmosphere With Microscale Geographic Predictors for Healthy Route Planning," *Advancing Earth and Space Science*, 2022.