Optimisation of Last Mile Delivery: Selecting the Best Warehouse Location using Simulation Models

Saket Abhaykumar Kulkarni
School of computing
National College of Ireland
Dublin, Ireland
x23102381@student.ncirl.ie

***Abstract*—Last Mile Delivery is always on the move to be optimised. This project aims to optimise the Last Mile delivery service for small townships by choosing the best possible Warehouse Location Provided by the Company We-Doo. A map of the town is generated and then algorithms like integer programming, A*, greedy and heuristic are used to find shortest routes to cover all customer locations. The best Warehouse location obtained from Monte-Carlo Optimisation is (5000,3740) on the map. The Cost base turned out to be the lowest and the deliveries quickest. This was concluded by using an ANOVA on different cost functions of sample simulations from other warehouse locations with respect to the chosen location.**

I. Introduction

The Optimisation process revolves around last mile delivery for the delivery company We-Doo who have managed to raise enough funding to to takeover the last mile delivery and look for an ideal warehouse location in small commuter towns to deliver their parcels which are dropped off in the warehouse location by the bigger logistics and these will intern be delivered to the customers by a local driver. The people of the town are also willing to pay a small fee for this service as it will make sure thier parcels are delivered on time. The problem arises when the delivery process to the customers has certain constrains and the deliveries need to happen only following those respectively. This is a well known problem called the Last Mile Problem (LMP) [1] which deals with transportation of goods from a public location to customers and can be rectified with Vehicle Routing Problem (VRP). In this case the public location will be the warehouse from where the parcels will be sent to delivery and customers remain as it is, where the transport method for the delivery is the delivery representative with an electric bike.

The following constraints are faced by the company to delivery the parcels:

- The cargo bike has a minimum range of 40 Km
- Average speed of 15 kmph
- The time for handing over the parcels to the customer consists of a call time (for the customer to come to the door) that follows an expovariate distributed on with a mean time of 40s and an additional handover time per parcel, which again follows an expovariate distribution with a mean time of 10s per parcel.
- The cumulative preparation time is said to be approximately 50s per parcel to be delivered
- The day end procedure is 10 minutes
- Operational cost of electricity cargo bike is 8c/km
- Drivers are paid 30 euro/hour with a minimum of 60 euro per working day.

The map of the town is randomly generated using the last four digits of the student ID as mentioned in the abstract. The dimensions of the township include 100 nodes and 60 customer locations with the centre of the circle being (5000,5000) and the radius of the circle is 4500 units. The seed value will be unchanged throughout to maintain the consistency and see what kind of challenges are faced. A total of 1Customer locations are set at random and are spread across the entire map generated by the author. A warehouse location is generated at random but also optimal at the same time. Algorithms used in identifying an ideal warehouse location and also to find the best possible route to follow to maximise deliveries and minimise the cost function within the given constraints and that has been done which will be seen in the sections going ahead. The routes were analysed using heuristic and greedy algorithms and different warehouse locations are tested to get the most optimal solution. The warehouse location is chosen using the monte-carlo optimisation to calculate the nearest best option and then chose the final warehouse location according to the algorithm.

II. Literature Review

A. Data - Driven Optimisation

Observing the previous work done on the similar problem, it is evident that the approach has always been tuned to try and get a more accurate solution to vehicle tracking and optimisation of deliveries whether it is food delivery or goods. Working on the last mile problem, the conclusion made in one of the studies was that data-driven optimisation techniques and prediction methods which are mainly machine learning models go well in order to bring out to solution to such real world problems out there. Data driven optimisation methods helped the the authors come up with a solution to be able to deliver food faster and with limiter drivers in their availability [2]. Previous approaches include statistics related to operations, sample average approximation(SAA) and robust optimisation. This study investigates last mile delivery problems using data driven methods, proposing an algorithm to handle actions between decision variables and uncertainty parameters.

B. Last Mile Problem

The Last Mile problem (LMP) involves in routing of vehicles efficiently in terms of delivery to the customers. The LMP model works with assigning delivery tasks to the vehicles which start routing from a central depot where the deliveries from the bigger logistics come in and then taken forward by the respective delivery person on the vehicle. This model helps in optimising the time and cost along with ensuring the deliveries on time according to the customers. All kinds of diverse options are considered with respect to number of drivers and vehicles being used to optimise the entire process and routing. The process involves network optimisation, defining number of nodes, solving the routing problems. The TSP model proposed in this paper incorporates various travel times which might be uncertain and are influenced by different factors and the objective is to minimise the travel time and costs while reaching the delivery demands [2]. The model aims to balance service quality and cost efficiency for the last mile operations.

C. Traveling Salesman Problem

Another study regarding last mile routing concluded that the proposed methodology for route optimisation and prediction showcased prominently (Ghosh et al., 2023). By using historical data and real time travel information, there is bridging of gaps between theoretical route optimisation and practical route prediction. This helps in solving the global travelling salesman problem very efficiently [3]. The work done by Jiang et al., (2019) [7] is focused on improvising the last mile delivery cost as well as reduction in carbon footprints created during the process of delivery and the model used is a variant of the travelling salesman problem called the iterative local search (ILS). This algorithms is used on scattered and clustered data which brought out positive results with respect to the objectives of the last mile problem.

D. Simulation of Last mile Logistics

Another study done by Bienzeisler and Friedrich (2023) focusing on transforming the last mile logistics by using a simulation based analysis dives into the topic of collection point delivery (CPD) which is the proposed solution to deliver the goods to a common point instead of individual addresses and that being more time consuming. MATSim is the framework used for the simulation analysis and it is a sophisticated agent-based transport simulation framework [6]. Following a CPD approach is similar to that of what this paper is also following, a single common warehouse and all the deliveries that come in there are then personally delivered to the final delivery locations, CPD could enhance the operational efficiency and even focus on the reduction of cost overall. The last mile delivery problem exists from a considerable amount of time and has been an objective of a lot of research in order to optimise it in terms of cost, sustainability and environmental footprint in terms of logistics. Finally for all these problems, CPD played a huge role in the transformation of the last mile delivery service.

E. Minimising Last Mile Delivery Cost

In the e-commerce world, last mile delivery plays the most important role in terms of service and cost as the company needs to make sure they don't go beyond budgets and also to keep the customer satisfaction high with on time delivery of the products. Due to the economic crisis and the costs of fuel going up, intern the operational costs also going up, last mile delivery has been critical in terms of expenditure. The routing system followed by the vehicle involved in the last mile delivery is extremely crucial as it depends on the cost for the vehicle used for delivery, basic transportation costs as well as cost for the labour. All these factors can be minimised by choosing an optimal route. A mathematical model was simulated using the SupplyChainGuru® modeling and simulation software. The study showed that the overall last- mile delivery cost is minimised by 22%. And this helped in reducing the number of vehicles used for delivery and by choosing a better optimal route. Failed delivery package count also decreases as the good routes cover most locations in a fixed time period making sure everything is delivered on time or even before time. Utilising the maximum possible capacity of vehicles while also increasing customer satisfaction by giving consumers a chance to select customer preferred time windows for package delivery which helps in routing according the preferred timing of the customer. This cluster-based delivery will improve the routing of the e-commerce logistic supply chain and will serve as a platform for extending the cluster-based delivery process to other industries as well [5].

III. Methodology

There are quite a few techniques involved in the entire process, the algorithms like integer programming , A*, Heuristic, Greedy algorithms for finding shortest path between two points and intern the shortest path throughout the whole map to cover the number of customers, simulation model to run simulations of deliveries per day for certain number of days, optimisation technique like the Monte-Carlo optimisation to try and find the best suitable warehouse location to make sure the deliveries are all smooth and remain in the constraints. The number of nodes in the map are a 100 and number of customer locations included are 60.

A. Generating Map Data

The map data is generated at random by passing the number of nodes and the number of customer locations. The euclidean distance between each point is calculated at random as well and also rounded off to the nearest 10th place to keep the values a whole number and the map generated will be more realistic in terms of a town. The edges generates will be the possible routes the delivery person will be following. The script for warehouse generation is written to allocate a warehouse location to one of the nodes which might or might not be suitable for the delivery process but the objective is to figure out which warehouse location would be best suited to optimise last mile delivery. The map data generated is then saved in pickled format to make sure its converted into a byte stream. It can be accessed easily by loading the pickled data and converting it back to an object as long as its in the same folder to ensure consistency while running algorithms on the data to find shortest path and other requirements. All the locations are set at random using the seed value of the last four digits of the student ID, in this case the seed value is 2381.

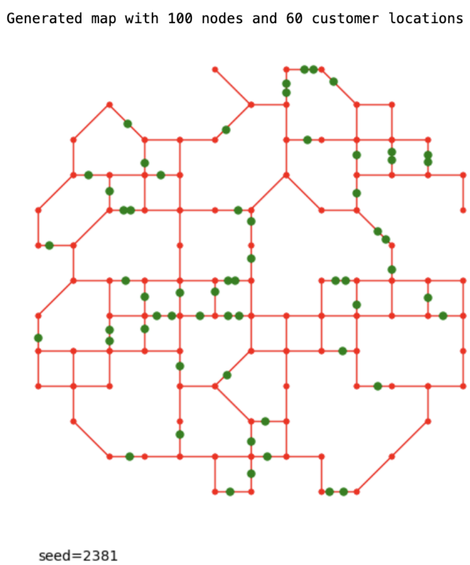

Fig. 1: Township map data generated including the customer location(green)

B. Generating Delivery Data

The map of the township is then used to generate a sample example of delivery data to simulate a simple process over a given number of input days, number of customers to deliver to and how many parcels will be delivered a day per person. The sample is again at random using the same seed value as mentioned above (2381). After running a demo example which consisted of 10 customers to be served, average 0.2 parcels to be delivered per person per day and over 5 days. The output received was a total of 16 parcels delivered.

Similarly, when the same program runs on the entire customer data, i.e all the customers are considered over a period of little over a month and with the same parameters of 0.2 parcels per day per customer, it is found that 424 parcels are delivered and all the days we well divided in terms of work rate and covering all parcels and customers.

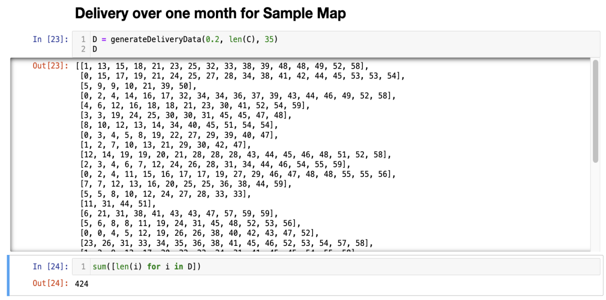

Fig. 2: Delivery simulation demo for 35 days

C. Finding Shortest Path

To find the shortest path and to implement best first search, the algorithm used firstly is the A* Algorithm [8]. It is one of the most preferred path finding algorithm due to its flexibility and wide range of contexts. It works in a greedy manner where it looks for the nearest point or node and chooses the most favourite node closest to the end point and keeps doing that until the final point is reached. The algorithm here takes the two point A and B and finds the shortest path where the edges and vertexes are the routes and locations.The function keeps a priority queue of the candidate shortest route considering the nearest points for each iterations and finally comes up with the most optimal path from start to end. In this algorithm, h(n) is the heuristic cost function and g(n) is the total cost of the path from start to the end.

$$f(n) = h(n) + g(n) \tag{1}$$

An example of the A* algorithm finding the shortest path between two points given as input is shown in figure 3.

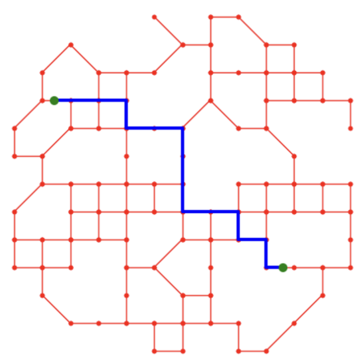

Fig. 3: Path finding example between two points A and B using A* Algorithm.

D. Finding Shortest Delivery Route

After finding the shortest path between two points it is now important to see the shortest path when it comes to delivering to all customers starting from the warehouse location to all the customer locations. An iterative comparison is done in order to see which of the algorithms are performing best and find the most optimal route. The algorithms in comparison are greedy and heuristic algorithms and a iterative integer programming algorithm to loop through the nodes. 3 loops are created, one for each algorithm and this is one on various maps using different warehouse locations and the path length is calculated using another function which helps in determining the shortest path and in terms of distance covered in meters. After using the different algorithms to loop through the map and find the shortest path, another optimisation method is used called the Monte-Carlo Optimisation. It is an optimisation method that uses random sampling to find the optimal solution to a problem, in this case the optimal solution the warehouse location from where the route is the shortest covering the number of customers. Random candidate solutions are generated and then iterations are run in order to come up with the final solution and that will be the final warehouse location [9]. The performance of the Monte-Carlo optimisation method on the township data is thoroughly discussed in the results and evaluation section.

The Algorithms used to find the shortest path are Mathematical algorithms, with the help of these algorithms, the shortest path was achieved.

Greedy Algorithm:

$$\text{Greedy Solution} = \underset{x \in X}{\operatorname{argmax}} f(x) \tag{2}$$

Heuristic Algorithm:

$$\text{Heuristic Solution} = \underset{x \in X}{\operatorname{argmin}} g(x) \tag{3}$$

Monte Carlo Algorithm:

$$\text{Monte Carlo Solution} = \frac{1}{N} \sum_{i=1}^{N} f(x_i) \tag{4}$$

Integer Programming:

$$\text{Integer Programming Solution} = \underset{\boldsymbol{x}}{\operatorname{argmin}} h(\boldsymbol{x}) \tag{5}$$

$$\text{subject to } \boldsymbol{x} \in \{0,1\}^n \tag{6}$$

E. Model Verification and Optimisation

To verify the model, a simulation is required to see how the delivery process goes by when it comes to delivering to a large number of customers over a longer period of time, for the simulation, the delivery data is generated again using the same function and then along with that a candidate warehouse location must also be given, in this case it can be the result obtained from the Monte-Carlo optimisation. There are 5 classes to consider in this process and each class being an important factor when it comes to the delivery. The 5 classes are :

1) Recorder : The recorder class keeps track of all the events like working time, costs and parcels during the delivery from the starting time till the ending until the driver is back to the warehouse after delivery.
2) Parcel : A very important factor to keep track off and the parcel class created in the script is responsible for checking whether the parcel left for delivery or not and if it was successfully delivered or had be brought back to the warehouse.
3) Customer : Conditions taken into consideration like whether the customer will be at home or not, waiting time includes the driver waiting till the customer opens and that is given a good 50 seconds.
4) Delivery center : The delivery center or the warehouse location being the most important objective of this project, handles all the functions related to the operations. When the parcel comes in and when it goes out for delivery is tracked along with any left over parcels for the day as well.
5) Driver : The driver class tracks the movement of driver from and to the warehouse while delivering the parcels in the evening.

F. Model Statistics and Cost Function

Statistics like the average working time per day, number of parcels left over in the warehouse and even the tour length taken every day is calculated in the recorder class with respect to each of the variables. It is important to track the numbers related to delivery everyday which can help in comparisons with routes taken previously, or even with choosing a warehouse locations. Which ever shows the best stats for example, the sample simulation with the left parcels left over after a day of delivery shows that the vehicle routing system is working well. ANOVA is used to test the significant difference between the cost functions. Again, the objective is to make sure the cost function is efficient and not to heavy which is the case in a lot of last mile delivery cases [5]. All the statistics will be discussed in detail in the results and evaluation section.

IV. Results and Interpretation

Initial and the main objective is to obtain the best location for the warehouse with the best vehicle routing process as well as cost efficiency and the following results that were achieved will be discussed and evaluated step by step. The chosen warehouse is the point (5000, 3740) on the map, it is not fully central or too far away from the centre which makes it the best option.

The optimisation runs multiple iterations using Monte-Carlo Algorithm from different candidate locations and then chooses the final one on the basis of the path length that was covered to deliver to all the 40 customers which was the sample in this case.

Going forward, the location is fixed to the point obtained by the algorithm and then compared to some other locations at random but the simulation models will run with the warehouse location at this point. As seen in figure 4, there are multiple iterations calculating the total distance of the route consisting

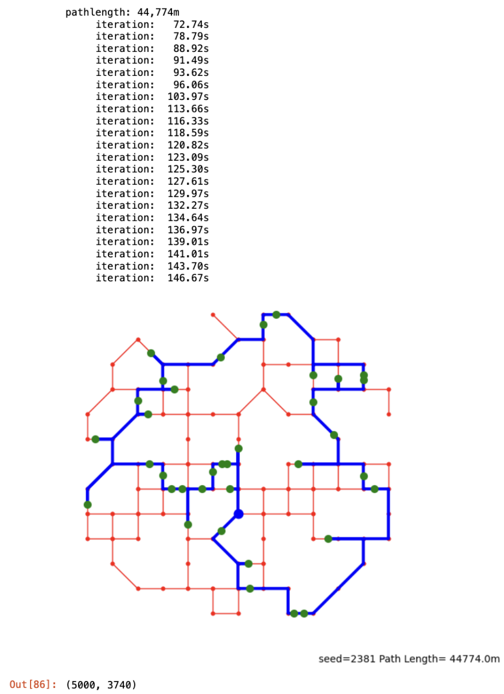

Fig. 4: Iterative Monte-Carlo optimisation with the final warehouse location

of 40 customers for the day and finally it comes down the point (5000, 3740) with the shortest path length of 44,774 meters which is approximately 44.77 km.

A. Simulation model

The SimPy framework in python used for event based simulation networks [11]. SimPy provides resources to model active components like customers and vehicles which in this case are the main factors involved in the process. The SimPy environment in this process followed for this project consists of the delivery data and the entire routing process which uses the delivery vehicle and the parcels are delivered to the customers. The simulation process used the same warehouse location obtained from Monte-Carlo and the sample routine is shown in figure 5. The simulation has been carried out for all the 60 customers over 200 days with an average of 0.15 parcels per person per day. The result obtained from the simulation was that 1855 parcels were delivered over that period of time. The model was also attempted on different candidate warehouse locations and the warehouse location chosen earlier gave the result sooner which makes it the best choice again.

B. working time

The working time plot shows some variation in terms of daily time and as seen in figure 6, There is not much variation in the average working time (red line in the graph). Over 200 days the working time is evenly distributed with an average of 124.16 which is perfect while considering the limitations in the working time.

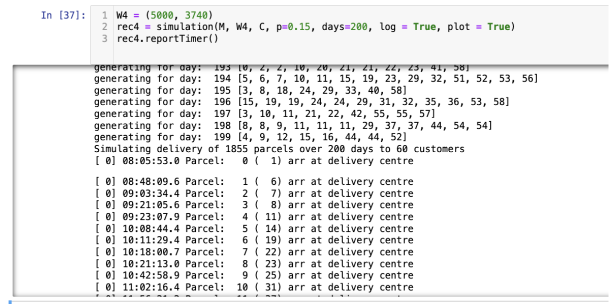

Fig. 5: Simulation result for the entire sample

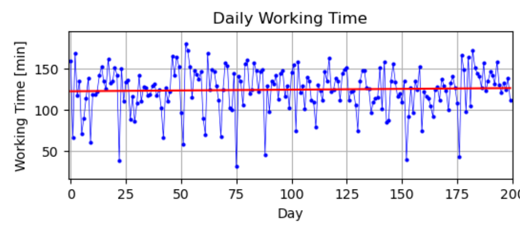

Fig. 6: Working time for delivery over 40 days

C. Route Length

The route length is the most important factor while optimising the process of delivery, the shortest route lengths are required to maintain efficiency on a daily basis. The route length is generated for each day in the script for a total of 200 days and as seen in figure 7 , there is variation on a day to day basis, some days its a little high and some days its a little low which is due to certain customer locations on the map that need to be covered. But overall the plot shows the consistency maintained each day to cover the right amount of distance and ensure quick delivery.

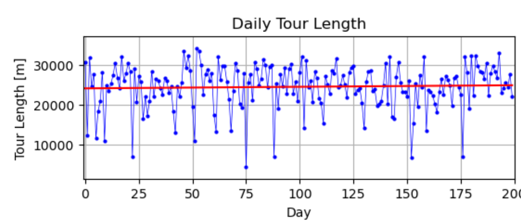

Fig. 7: Daily tour length for delivery over 40 days

D. Parcels Left Over

With respect to parcels an cost of the operation, both are better if the numbers are lower. After the simulation process with the same parameters and the same warehouse location, the number of parcels left at the end of the simulation is zero but in between the days there were a few left over for the next days and all were delivered in priority. Therefore the inventory on the last day is empty with no left over parcels.

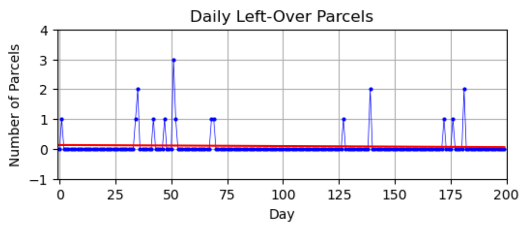

Fig. 8: Plot for Left over Parcels

E. Cost Functions

The daily cost for a driver must be a minimum of 60 euro with hourly wage of 30 euro. The range of the electric bike is 40km and as seen above, none of the days cross that mark which means the cost can be calculated going further. The cost for daily operations which is 0.08c/km is also calculated in the recorder and is used in the determination of the best warehouse location. A sample simulation is run of 10 different simulations for 200 days each and the one of the warehouse location is the chosen one. After Comparing the costs from each of the warehouse locations, An ANOVA can be used to test the variance in the costs and to see if there a significant difference among the means of the costs. The objective here again is to chose the warehouse to location that returns the lowest cost base, therefore now that the warehouse location has been chosen, proof of low cost base must be provided.

After running the simulation for 200 days, the average cost per day for the driver obtained from the program is 62.08 which is low number and exactly close to how much the driver needs to be paid on a minimum per day. The cost function for the daily operations cost comes up to 0.08c/km. The results are as shown in the figure 9b which is around 1.96 euro per day on average with the delivery vehicle being an electric bike, a lot of costs are saved and also makes the process more sustainable.

```
In [40]:  #daily driver cost average
          rec.driverCost().mean()

Out[40]: 62.08
```

(a) Average cost of driver per day

```
In [39]:  rec.dailyOperationsCost().mean()

Out[39]: 1.9660599999999997
```

(b) Average Daily Operations Cost

Fig. 9: Average Cost for Daily Driver and Operations

1) Statistical Significance of the Cost Functions: One of the most important factor influencing the choice of the warehouse location is the cost, the main cost in this situation are discussed on the above section where the average driver cost and the operational costs are discussed. Since both show a reasonable number in terms of the cost base for the warehouse, their significance must be identified. To identify the significance, the sample 5 is the simulation with the main warehouse location (5000, 3740). After carefully analysing all the daily costs of the simulation samples from different warehouse locations, the daily costs for the main location of sample 5 turned out to be the lowest and most economical in the long run.

The results of the ANOVA test also show that there is significant difference between the costs for both the functions and are shown in the figure 10a and figure 10b as follows. The results can conclude for now that the cost on a daily basis is very different for different warehouse locations and the current costs for the parameters given is lowest at the point given by Monte Carlo which proves the power of the algorithm and how well it behaves when it comes to decision making as well as finding the best possible outcome without any sort of solid input. The analysis of the variance also proves that the cost base was the lowest for the main location in sample 5.

```
F-statistic: 2.5953968149646336
P-value: 0.005631236794826016
ANOVA indicates a significant difference among the driver costs (p < 0.05)
```

(a) Anova for Daily Driver Cost

```
F-statistic for operational costs: 3.818372920386979
P-value for operational costs: 8.469275735550005e-05
ANOVA indicates a significant difference among operation costs (p < 0.05)
```

(b) Anova for Daily Operations Cost

Fig. 10: ANOVA Results from both the Cost Functions

V. Reflection and Future Work

Reflecting on the matter, The entire model seemed to perform considerably well and all the results achieved met the requirements, A more precise model going a little deeper into the problem would do much better in terms of accuracy and time. The model was used for a small township and that was due to the computational power of the system used and the lack of a GPU in the system due to which the program would even stop running at times and has to be restarted and run from scratch. The current parameters used ran successfully and took reasonable time. In real life situations when working with bigger towns, it will require more precision and higher power to run the simulation programs. Overall, the results achieved were good but it can get better which requires more study in the operations research field and while optimising the deliveries, more factors can be considered while deciding the best location for the warehouse such as cost of the rent for the warehouse, customer satisfaction, weather, weekends and holidays etc. These are some examples and on a bigger project, these can be applied to get the best possible location and optimise Last Mile Delivery.

References

[1] Wang H, Odoni A (2014) *Approximating the performance of a "last mile" transportation system.* Transport Sci 50(2):659–675

[2] H. Chu, W. Zhang, P. Bai, and Y. Chen, *"Data-driven optimization for last-mile delivery,"* Complex & Intelligent Systems, vol. 9, Feb. 2021, doi: https://doi.org/10.1007/s40747-021-00293-1.

[3] Ghosh, M., Kuiper, A., Mahes, R., and Maragno, D., *"Learn global and optimize local: A data-driven methodology for last-mile routing,"* Computers & Operations Research, vol. 159, pp. 106312–106312, Nov. 2023, doi: https://doi.org/10.1016/j.cor.2023.106312.

[4] J. K. Chowlur Revanna and N. Y. B.Al-Nakash, *"Vehicle Routing Problem with Time Window Constrain using KMeans Clustering to Obtain the Closest Customer,"* Global Journal of Computer Science and Technology, pp. 49–64, Jan. 2022, doi: https://doi.org/10.34257/gjcstdvol22is1pg49.

[5] G. Kasuri Abhilashani, M. I. D. Ranathunga and A. N. Wijayanayake, *"Minimising Last-Mile Delivery Cost and Vehicle Usage Through an Optimised Delivery Network Considering Customer-Preferred Time Windows,"* 2023 International Research Conference on Smart Computing and Systems Engineering (SCSE), Kelaniya, Sri Lanka, 2023, pp. 1-7, doi: 10.1109/SCSE59836.2023.10215012.

[6] L. Bienzeisler and B. Friedrich, *"Transforming Last-Mile Logistics: A Simulation-Based Analysis of Collection Point Delivery Strategies,"* 2023 IEEE 26th International Conference on Intelligent Transportation Systems (ITSC), Bilbao, Spain, 2023, pp. 2711-2718, doi: 10.1109/ITSC57777.2023.10422582.

[7] L. Jiang, H. Chang, S. Zhao, J. Dong and W. Lu, *"A Travelling Salesman Problem With Carbon Emission Reduction in the Last Mile Delivery,"* in IEEE Access, vol. 7, pp. 61620-61627, 2019, doi: 10.1109/ACCESS.2019.2915634.

[8] *"Introduction to A*,"* theory.stanford.edu. https://theory.stanford.edu/ amitp/GameProgramming/AStarComparison.html

[9] T. Homem-de-Mello and G. Bayraksan, *"Monte Carlo sampling-based methods for stochastic optimization,"* Surveys in Operations Research and Management Science, vol. 19, no. 1, pp. 56–85, Jan. 2014, doi: https://doi.org/10.1016/j.sorms.2014.05.001.

[10] *"Heuristic algorithms - Cornell University Computational Optimization Open Textbook - Optimization Wiki,"* optimization.cbe.cornell.edu. https://optimization.cbe.cornell.edu/index.php?title=Heuristic algorithms

[11] Overview - SimPy 4.1.1 documentation. Available at: https://simpy.readthedocs.io/en/latest/index.html (Accessed: 08 April 2024).

[12] A. Gupta, S. Ghosh, and A. Dhara, *"Deep Reinforcement Learning Algorithm for Fast Solutions to Vehicle Routing Problem with Time-Windows,"* in ACM International Conference Proceeding Series, Jan. 2022, pp. 236–240. doi: 10.1145/3493700.3493723.

[13] J. Keerthy, C. Revanna, N. Yousif, and B. Al-Nakash, *"Vehicle Routing Problem with Time Window Constrain using KMeans Clustering to Obtain the Closest Customer,"* 2022, [Online]. Available: https://www.cogoport.com/blogs/transport-

[14] N. Boysen, S. Fedtke, and S. Schwerdfeger, "Last-mile delivery concepts: a survey from an operational research perspective," OR Spectrum, vol. 43, no. 1, pp. 1–58, Mar. 2021.