

DIGITAL

A Two-Stage Clustering and Genetic Algorithm Approach for Courier Delivery Clustering and Route Optimization

Evan Widyadhana Anargya¹, Maulin Masyito Putri^{1*}, Linggarwangi Yuniar¹

¹Department of Logistics Engineering, Universitas Internasional Semen Indonesia, Gresik, Indonesia

*Corresponding author: maulin.putri@uisi.ac.id

ABSTRACT

Efficient delivery area allocation and route planning are essential for improving operational performance in last-mile logistics. However, inappropriate delivery area grouping can lead to longer travel distances and reduced delivery efficiency. This study proposes a two-stage approach for courier delivery optimization consisting of delivery area clustering and route optimization. In the first stage, delivery areas were clustered using Single Linkage and K-Means based on geographical proximity. In the second stage, delivery routes were optimized using a Genetic Algorithm (GA) within the Capacitated Vehicle Routing Problem (CVRP) framework while considering vehicle capacity constraints. The proposed approach was evaluated using delivery demand data from a courier service hub and compared with the existing delivery system. The results indicate that the K-Means clustering approach produced the shortest total travel distance among the evaluated scenarios. Demand analysis showed that the highest weekend demand occurred on Sundays with 199 packages, while the highest weekday demand occurred on Wednesdays with 147 packages. Compared with the existing routing system, the K-Means-based approach reduced total travel distance by 37.6% on Sundays and 5.51% on Wednesdays. These findings demonstrate that the proposed cluster-first, route-second approach can effectively reduce travel distance and improve delivery efficiency in last-mile logistics operations.

Keywords: Vehicle Routing Problem, Clustering, Genetic Algorithm, Last-Mile Delivery.

1. INTRODUCTION

The rapid growth of e-commerce has substantially increased parcel delivery demand, making last-mile delivery one of the most critical and costly stages of the logistics supply chain. Increasing customer expectations for fast, reliable, and flexible delivery services have forced logistics providers to continuously improve operational efficiency while maintaining service quality and controlling transportation costs. In Indonesia, logistics companies are responsible for planning, supervising, and controlling the movement of goods from the point of origin to the final destination (Amirah Kamila Safitri et al., 2023). As delivery volumes continue to rise,

effective planning of delivery operations has become increasingly important to ensure sustainable and cost-efficient last-mile logistics.

One of the key challenges in last-mile delivery is the allocation of delivery service areas. Inefficient assignment of customers into delivery zones may create geographically dispersed service regions, leading to longer travel distances, increased delivery time, unbalanced courier workloads, and higher operational costs. These challenges become even more significant when delivery vehicles operate under capacity constraints because poor delivery area allocation directly affects the quality of subsequent route planning. Consequently, delivery area planning and vehicle routing should be considered as interdependent decisions rather than separate optimization problems.

To improve last-mile delivery performance, previous studies have investigated various approaches for delivery region planning and vehicle routing optimization. Recent studies have highlighted that efficient partitioning of delivery regions plays an important role in improving routing performance because the quality of customer clustering directly influences travel distance, workload balance, and overall delivery efficiency. Leung et al. (2026a) proposed a weighted self-organizing map for delivery region partitioning and demonstrated that generating compact delivery regions significantly improves routing efficiency compared with conventional clustering techniques. Similarly, Sun & Arslan (2025) emphasized that incorporating operational knowledge into routing decisions can substantially improve the quality of last-mile delivery routes. In addition, numerous studies have shown that metaheuristic algorithms remain effective for solving large-scale Vehicle Routing Problems (VRPs) because they are capable of producing high-quality solutions within acceptable computational time (Braekers et al., 2016; Vidal, 2022).

Despite these advances, several research gaps remain. First, many previous studies primarily focus on optimizing vehicle routes after delivery regions have already been established, providing limited investigation into how the quality of delivery area allocation affects routing performance. Second, although integrated cluster-first, route-second approaches have received increasing attention, most studies either develop new clustering algorithms or adopt a single clustering technique without comparatively evaluating conventional clustering methods for delivery area formation. Consequently, there is still limited empirical evidence regarding the effectiveness of different clustering approaches in supporting subsequent route optimization. Third, relatively few studies validate integrated clustering-routing frameworks using real operational data from courier services while simultaneously considering vehicle capacity constraints. These limitations indicate the need for a practical framework that systematically integrates delivery area allocation and route optimization for real-world courier operations.

To address these research gaps, this study proposes a two-stage cluster-first, route-second framework for courier delivery optimization. Unlike previous studies that employ a single clustering strategy or focus solely on routing optimization, this research compares two widely used clustering approaches, namely Single Linkage and K-Means, to evaluate their effectiveness in generating delivery areas that support more efficient routing. The clustered delivery regions are subsequently optimized through the Capacitated Vehicle Routing Problem (CVRP) using a Genetic Algorithm while considering vehicle capacity constraints. The proposed framework is validated using real delivery demand data obtained from a courier

service hub and compared with the company's existing delivery practice. The findings are expected to provide practical guidance for logistics companies in improving last-mile delivery efficiency while contributing empirical evidence on the relationship between delivery area formation and routing performance.

2. METHODS

2.1. Research Design

This study adopts an applied quantitative research design using operational delivery data collected from a courier service hub in Surabaya, Indonesia. The hub processes approximately 150 parcels per day using six motorcycle couriers (Table 1). Customer information includes geographic coordinates (latitude and longitude) and parcel demand, which serve as the primary inputs for the analysis. The research consists of two sequential analytical stages. First, customer locations are grouped into delivery areas using Single Linkage and K-Means clustering based on geographic proximity. The resulting clusters are then evaluated to determine their suitability for delivery operations. Second, the delivery routes within each cluster are optimized using the CVRP, which incorporates motorcycle carrying capacity as the primary operational constraint. The optimization process is solved using a GA to determine the delivery sequence that minimizes the total travel distance. The performance of each clustering scenario is evaluated by comparing the resulting travel distance with the existing delivery operation. The comparison provides an assessment of the effectiveness of each clustering method in supporting efficient last-mile delivery.

Table 1. Existing courier delivery distances

Courier	Coverage Area	Distance (km)
1	Lidah Kulon & Sumur Welut	7.2
2	Bangkingan & Jajar Tunggal	14
3	Wiyung	2.9
4	Babatan & Lidah Wetan	6.4
5	Balas Klumprik	5.2
6	Jeruk	4.5

All data were collected directly from Hub through semi-structured interviews with the hub manager and direct field observation (Table 2). Five operational data categories were recorded:

1. Depot coordinates: GPS latitude and longitude of the Hub (depot), recorded using Google Maps.

2. Delivery-point coordinates: GPS latitude and longitude of each parcel-recipient address, recorded using Google Maps; distances are treated as symmetric.
3. Daily parcel demand: the number of parcels (*demand*) to be delivered at each address per observation day.
4. Number of active couriers: six motorcycle-riding couriers operating from the hub.
5. Vehicle capacity constraints: maximum load Q per motorcycle 30 small packages, 25 mixed packages, or 20 large packages.

Table 2. Respondent Data

Point	Code	Demand	Latitude (ϕ)	Longitude (λ)
Depot	0	0	-7.3140	112.6677
Customer 1	1	2	-7.3140	112.6698
Customer 2	2	3	-7.3257	112.6895
Customer 3	3	1	-7.3255	112.6884
Customer 4	4	2	-7.3249	112.6884
Customer 5	5	3	-7.3258	112.6867
Customer 6	6	1	-7.3262	112.6864

2.2. Spatial Distance Metrics

Before clustering, the pairwise distance matrix between all nodes (depot and delivery points) must be computed. Two distance formulations are applied in this study:

a. Haversine Distance

The clustering process is carried out by calculating the polar angle of each customer in radians, which is then converted into degrees to facilitate the sorting process; customer points are subsequently grouped into clusters based on the order of these polar angles until the vehicle capacity is reached. Once the clusters are formed, the Nearest Neighbor algorithm is applied to arrange the visiting sequence within each cluster, starting and ending at the depot, so that the total travel distance can be calculated for each resulting route. By combining the Haversine distance with the angle-based clustering algorithm, this study aims to improve the efficiency of distribution route optimization by providing a more accurate representation of geographical distances between customer locations, thereby generating delivery clusters that better reflect real-world distribution conditions (Boyacı et al., 2021). The Haversine formula calculates the great-circle distance between two geographic coordinates, expressed by latitude (ϕ) and longitude (λ), on the Earth's surface with an average radius of 6,371 km. Compared with planar distance measures, the Haversine distance provides a more realistic approximation of spatial separation for geographically distributed locations and is therefore widely adopted in transportation, logistics, and vehicle routing applications (Ballou et al., 2002; Boyacı et al., 2021). The Haversine distance is calculated as follows:

$$x = (\lambda_j - \lambda_i) \cdot \cos\left(\frac{(\varphi_i + \varphi_j)}{2}\right) \quad (1)$$

$$y = \varphi_j - \varphi_i \quad (2)$$

$$d_{ij} = \sqrt{x^2 + y^2} \cdot R \quad (3)$$

where 1 degree = 0.0174532925 radians. Applying Equations (1)–(3) to Table 2 produces the 7×7 Haversine distance matrix (km) used in Section 2.4.

b. Euclidean Distance

Compares two distance measurement methods, the Euclidean method and the Haversine formula, for determining the nearest public facilities. The Euclidean method is a heuristic function that calculates the straight-line distance between two points without considering obstacles, similar to computing the length of a diagonal line in a triangle, and therefore does not account for the curved shape of the Earth (Maria et al., 2020). In contrast, the Haversine formula is used to determine the arc distance between two points based on latitude and longitude coordinates, making it more suitable for geographic data on the Earth's spherical surface (Table 3). When coordinates are reprojected onto a Cartesian plane (x, y), the Euclidean distance is applied equation 4.

$$d_{ij} = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2} \quad (4)$$

In this study, Haversine (Eq. 1–3) is the primary distance metric because all coordinate data are in geographic (latitude/longitude) form. The Euclidean formula is noted as an alternative applicable after coordinate projection.

Table 3. Haversine distance matrix (km)

	0	1	2	3	4	5	6
0	0	0.233	2.729	2.617	2.581	2.475	2.465
1	0.233	0	2.527	2.417	2.379	2.281	2.275
2	2.729	2.527	0	0.118	0.150	0.300	0.345
3	2.617	2.417	0.118	0	0.072	0.186	0.238
4	2.581	2.379	0.150	0.072	0	0.209	0.269
5	2.475	2.281	0.300	0.186	0.209	0	0.063
6	2.465	2.275	0.345	0.238	0.269	0.063	0

2.4 K-Means Clustering Model

The K-Means algorithm was employed to partition customer locations into geographically compact delivery zones based on their spatial coordinates. The clustering process begins by selecting an initial set of centroids corresponding to the predetermined number of clusters (k).

Subsequently, the distance between each customer and the centroids is calculated, and each customer is assigned to the nearest centroid. The centroids are iteratively updated by computing the mean position of all points within each cluster until the cluster assignments converge, resulting in compact and homogeneous customer groups (Lloyd, 1982; Nie et al., 2021). Spatial partitioning of delivery zones using clustering techniques has been shown to improve delivery efficiency by reducing geographical dispersion, balancing courier workloads, and facilitating more effective vehicle routing (Leung et al., 2026b). The objective of the K-Means algorithm is to minimize the within-cluster sum of squared distances, as expressed in the following objective function. The operational algorithm follows these exact sequential steps:

- 1 Prepare and input the multi-dimensional dataset containing the latitude and longitude coordinates for each customer location.
- 2 Initialize K random centroids within the spatial bounding box corresponding to the pre-determined number of clusters.
- 3 Compute the great-circle distance from every individual customer coordinate to all initialized centroids using the Haversine metric.
- 4 Assign each customer node to its nearest cluster centroid based on absolute minimum spatial distance.
- 5 Evaluate convergence criteria by checking if cluster re-assignments or centroid adjustments fall below the termination threshold.
- 6 Recalculate the position of each centroid by taking the arithmetic mean of the coordinates of all nodes assigned to that cluster.
- 7 Store final converged cluster membership variables and node assignments.
- 8 Generate and export visual thematic maps of the finalized courier delivery zones.

As an illustrative baseline of the algorithm's performance, the clustering execution for Saturday successfully partitioned the service area into 6 distinct courier zones, encompassing 140 unique customer nodes with an aggregate demand of 177 packages. The numerical distributions per cluster zone are structured in Table 4.

Table 4. K-Means Clustering Layout

Courier / Zone	Number of Customer Points	Total Demand (Packages)
1	18	26
2	6	9
3	27	28
4	10	16
5	53	61
6	26	37
Total	140	177

2.5 Genetic Algorithm

The offspring chromosomes subsequently undergo mutation with a mutation probability (P_m) of 0.05 to maintain population diversity and reduce the risk of premature convergence. The mutated offspring are then incorporated into the next generation, and the evolutionary process of selection, crossover, and mutation is repeated iteratively until the stopping criterion is satisfied or the best fitness value converges, resulting in a near-optimal travel distance solution for each vehicle (Ou et al., 2024; Wouda et al., 2024). To solve the NP-hard CVRP within each clustered delivery zone, a customized GA was implemented. The algorithm employed a population size of 30 chromosomes, a maximum of 1000 generations, a crossover probability (P_c) of 0.80, and a mutation probability (P_m) of 0.05. These parameter settings were selected to provide an appropriate balance between exploration and exploitation while maintaining computational efficiency and solution quality for large-scale combinatorial optimization problems (Ou et al., 2024; Wouda et al., 2024). The iterative routing optimization routine proceeds through the following computational workflow:

- 1 Input operational datasets, specifically localized customer demand arrays, vehicle carrying capacity constraints ($Q = 30$), and fleet size configurations.
- 2 Initialize control hyper-parameters: population size, crossover rate, mutation rate, maximum generation limits, chromosome encoding length, and available fleet sizes.
- 3 Generate the initial generation population consisting of randomly ordered sequence chromosomes matching the specified population dimension.
- 4 Evaluate individual chromosomes against rigid CVRP capacity boundaries, inserting depot return sequences whenever cumulative demand exceeds the vehicle payload limit.
- 5 Map vehicle routing trajectories matching the structural gene order of each chromosome from the initial population bank.
- 6 Compute individual fitness values, defined mathematically as the inverse of the total cumulative routing distance, to enforce minimization objectives.
- 7 Execute parent selection using a Roulette Wheel Selection mechanism, where selection probability scales proportionally with high fitness metrics.
- 8 Perform crossover operations on selected parental pairs according to the fixed crossover probability to exchange route sequences.
- 9 Apply genetic mutation to offspring chromosomes according to the mutation probability to prevent premature local convergence.
- 10 Assemble the successor population pool from the mutated offspring, evaluate generational fitness, and track the best performing chromosome across iterations.
- 11 Isolate and export the chromosome possessing the absolute highest fitness value once the maximum generation limit is met, rendering the globally optimized physical routing plan.

For Saturday's operational log, the integrated K-Means-GA framework produced a total routing distance of 64.30 km. This reflects an improvement over the alternative Single Linkage-GA pipeline, which required a longer distance of 69.72 km (Table 5). To demonstrate generalized empirical validity across the full longitudinal study period, Table 5 provides a comprehensive summary of total cumulative travel distances for both algorithmic pipelines.

The experimental metrics compiled in Table 5 demonstrate that the K-Means-GA routing model consistently achieves shorter overall travel distances on almost all observation days compared to the Single Linkage-GA alternative. Consequently, K-Means clustering serves as the more reliable spatial foundation for determining optimal, resource-efficient courier routes within this logistical network.

Table 5. Total Cumulative Travel Distance (km) Performance Comparison

Operational Day	Single Linkage-GA	K-Means-GA
Saturday (Day 1)	69.72	64.30
Sunday	80.36	79.33
Monday	51.84	54.59
Wednesday	66.56	57.76
Thursday	62.69	57.34
Friday	60.08	52.61
Saturday (Day 8)	62.34	59.42

3. RESULTS AND DISCUSSIONS

3.1 Courier Zone Classification

Parcel demand collected from hub over the seven-day observation period fluctuates considerably, with Sunday recording the highest volume (199 parcels) and Wednesday the lowest (147 parcels), as summarized in Table 6. This demand variability motivates a flexible, recomputed clustering scheme rather than a single fixed zone boundary.

Table 6. Daily Parcel Demand

Day	Sat.	Sun.	Mon.	Wed.	Thu.	Fri.	Sat.(28)
Demand (parcels)	177	199	–	147	–	–	–

Table 7. Cluster Membership Comparison Single Linkage vs. K-Means

Cluster	SL Demand (pcs)	SL Members	K-Means Demand (pcs)	K-Means Members
1	57	45 pts	26	18 pts
2	51	34 pts	9	6 pts
3	12	9 pts	28	27 pts
4	49	45 pts	16	9 pts
5	6	5 pts	61	31 pts
6	2	1 pt	37	26 pts
Total	177	139 pts	177	117 pts

Minor discrepancy from 140 reflects shared boundary points cross-checked manually against the MATLAB cluster plot. Table 7 shows that Single Linkage produces highly unequal clusters Cluster 6 contains only a single point (2 parcels) while Cluster 1 carries 57 parcels a direct consequence of the chaining effect, in which each merge step attaches the nearest unassigned point regardless of overall zone balance. K-Means, which iteratively minimizes within-cluster squared distance to a centroid, distributes demand more evenly (9–61 parcels per cluster) and keeps point counts more proportional to cluster demand, indicating more geographically compact, centroid-balanced zones.



Figure 1. Comparison of Single Linkage and K-Means Clustering Results

3.2 Delivery Route Optimization Using GA under CVRP Constraints

Within each cluster, the optimal visiting sequence was determined by Ga formulated as a CVRP. Table 8 lists the GA parameters used in the full-scale MATLAB implementation, together with the three capacity tiers applied according to parcel size.

Table 8. GA Parameters and Capacity Tiers

Parameter	Value	Description
Population size	30	Number of chromosomes per generation
Maximum generations	1,000	Stopping criterion

Crossover probability (p_c)	0.8	Order Crossover (OX)
Mutation probability (p_m)	0.05	Swap Mutation
Capacity Q small parcels	30 pcs	Highest capacity tier
Capacity Q mixed parcels	25 pcs	Medium capacity tier
Capacity Q large parcels	20 pcs	Lowest capacity tier

A chromosome encodes a permutation of delivery nodes, beginning and ending at the depot (node 0). Any chromosome whose cumulative demand along the route exceeds the relevant Q triggers a penalty in fitness evaluation, lowering its selection probability. Table 9 illustrates the fitness evaluation mechanism $F(i) = Q(i)/\sum Q(i)$, where $Q(i) = 1/Z(i)$, using the validated 7-point dummy population after Roulette Wheel Selection, Order Crossover, and Swap Mutation.

Chromosome k1 yields the lowest distance (5.606 km) and the highest fitness (0.254), and was therefore selected as the manually-verified best route; the MATLAB simulation independently converged to route 0-1-5-6-2-4-3-0 at 5.760 km, confirming agreement between the manual and software computations. Table 10 presents the resulting Saturday delivery routes per courier for both clustering inputs, after the GA was applied separately to each cluster.

Table 9. Fitness Evaluation of the Post-Mutation Chromosome Population

Chr.	Route Sequence	Z (km)	Q(i)	F(i)
k1	0-1-6-5-3-2-4-0	5.606	0.178	0.254
k2	0-1-5-6-2-4-3-0	5.760	0.174	0.248
k3	0-6-2-3-4-5-1-0	5.724	0.175	0.249
k4	0-5-4-3-2-6-1-0	5.727	0.175	0.249

Table 10. GA Route Results per Courier

Courier	SL-GA Distance (km)	SL Demand	K-Means-GA Distance (km)	K- Means Demand	Better Input
1	9.69	57	7.64	26	K-Means
2	15.50	51	3.74	9	K-Means
3	7.11	12	8.93	28	Single Linkage
4	25.14	49	4.28	16	K-Means
5	9.20	6	34.94	61	Single Linkage
6	3.08	2	4.77	37	Single Linkage
Total	69.72	177	64.30	177	K-Means (overall)

Distances are GA-optimized route lengths per courier; demand columns show total parcels assigned to that courier under each clustering method. Although Single Linkage produces a shorter individual route for three of six couriers (3, 5, and 6), this advantage arises mainly because those particular clusters happened to receive fewer, more tightly grouped points by chance, not because Single Linkage systematically clusters better. At the aggregate level, K-Means-GA still achieves the lower total distance (64.30 km vs. 69.72 km), because the courier-

2 and courier-4 routes under Single Linkage are disproportionately long (15.50 km and 25.14 km), reflecting the elongated, chained cluster shapes discussed in Section 3.1.

3.3 Model Performance Evaluation and Comparative Discussion

Table 11 compares the total weekly travel distance obtained using the Single Linkage-GA and K-Means-GA models.

Table 11. Total Travel Distance Comparison

Day	SL-GA (km)	K-Means-GA (km)	Difference (km)	K-Means Advantage
Saturday	69.72	64.30	-5.42	-7.8%
Sunday	80.36	79.33	-1.03	-1.3%
Monday	51.84	54.59	+2.75	+5.3%
Wednesday	66.56	57.76	-8.80	-13.2%
Thursday	62.69	57.34	-5.35	-8.5%
Friday	60.08	52.61	-7.47	-12.4%
Saturday (28th)	62.34	59.42	-2.92	-4.7%
Average	64.80	60.76	-4.03	-6.2%

K-Means-GA outperforms Single Linkage-GA on six of the seven observed days, achieving daily travel distance reductions ranging from 1.3% (Sunday) to 13.2% (Wednesday), with an average weekly reduction of 6.2%. This result is consistent with the spatial compactness advantage discussed in Section 3.1. Centroid-based partitioning generates more compact and balanced delivery zones, whereas the chaining effect inherent in Single Linkage often produces elongated clusters that require the GA to construct longer connecting routes, regardless of the routing parameters employed. The only exception occurs on Monday, when Single Linkage-GA achieves a 5.3% shorter travel distance. This reflects a known limitation of K-Means, whose solution depends on the initial centroid placement and may occasionally converge to a locally suboptimal partition for a particular spatial distribution of delivery points.

These findings are consistent with recent studies highlighting the importance of integrating customer clustering with vehicle routing in last-mile logistics. Effective spatial partitioning prior to route optimization has been shown to generate more compact delivery territories, reduce travel distance, improve workload balance, and enhance overall routing efficiency. In particular, cluster-first, route-second strategies enable routing algorithms to search within geographically coherent customer groups, thereby producing shorter and more practical delivery routes than routing methods applied directly to unpartitioned customer sets (Dupas et al., 2024; Ramirez-Villamil et al., 2023). These observations further support the superior performance of the K-Means-GA framework, where compact centroid-based clusters provide a more favorable search space for the Genetic Algorithm than the elongated clusters produced by Single Linkage.

From a practical perspective, the average 6.2% reduction in weekly travel distance shortens courier travel time, decreases fuel consumption, and reduces the route overlap observed during field observations at the hub, where different couriers frequently traversed the same streets in opposite directions. By eliminating these redundant travel segments, the proposed approach improves route efficiency and aligns fuel expenditure more closely with the fixed delivery compensation received by each courier, thereby addressing the operational imbalance identified in Section 2.1.

Table 12. Summary of Key Findings

Aspect	Finding
Cluster balance	K-Means more evenly distributes demand (9-61 pcs) vs. Single Linkage chaining (2-57 pcs)
Best single chromosome	k1: 0-1-6-5-3-2-4-0, $Z = 5.606$ km, $F(i) = 0.254$
Weekly average distance	K-Means & GA: 60.76 km vs. SL & GA: 64.80 km
Average improvement	6.2% shorter total distance using K-Means & GA
Exception day	Monday - Single Linkage outperforms K-Means by 5.3%

These results support adopting the K-Means-based Cluster-First, Route-Second model as the preferred improvement scenario for the hub, while the Monday exception suggests that day-specific re-initialization of centroids, or a hybrid strategy combining hierarchical pre-processing with centroid-based balance, may further improve the consistency of distance reduction across all delivery days.

4. CONCLUSION

This study proposed a two-stage approach for courier delivery optimization by integrating delivery area clustering with route optimization. In the first stage, delivery locations were grouped using Single Linkage and K-Means clustering based on geographical proximity, while in the second stage, delivery routes were optimized using a GA within the CVRP framework. The proposed cluster-first, route-second approach provides a systematic method for improving both delivery area allocation and route planning in last-mile logistics.

The experimental results show that the K-Means-based approach consistently generated more balanced delivery clusters and achieved the shortest total travel distance among the evaluated scenarios. Compared with the existing delivery system, the proposed method reduced total travel distance by up to 37.6% on Sundays and 5.51% on Wednesdays, demonstrating its effectiveness in improving delivery efficiency while considering vehicle capacity constraints. In addition, demand analysis revealed that delivery volume varied considerably across operating days, with the highest demand occurring on Sundays and the highest weekday demand on Wednesdays.

Overall, the findings indicate that integrating K-Means clustering with a GA offers an effective decision-support approach for courier delivery clustering and route optimization. The proposed framework can assist logistics service providers in designing more efficient delivery

areas, reducing travel distance, and improving operational efficiency in last-mile delivery operations. A

Acknowledgments

The authors wish to thank the reviewers for their valuable suggestions to improve the quality of this paper. We also express our gratitude to the Logistics Engineering Department of Universitas Internasional Semen Indonesia for the financial support.

5. REFERENCES

- Amirah Kamila Safitri, Mohamad Rizan, & Rahmi. (2023). Peranan E-Service Quality, Rates Dan Customer Trust Bagi Repurchase Intention Jasa Kurir. *Jurnal Ekonomi*, 28(2), 234–254. <https://doi.org/10.24912/je.v28i2.1646>
- Ballou, R. H., Rahardja, H., & Sakai, N. (2002). Selected country circuitry factors for road travel distance estimation. *Transportation Research Part A: Policy and Practice*, 36(9), 843–848. [https://doi.org/10.1016/S0965-8564\(01\)00044-1](https://doi.org/10.1016/S0965-8564(01)00044-1)
- Boyacı, B., Dang, T. H., & Letchford, A. N. (2021). Vehicle routing on road networks: How good is Euclidean approximation? *Computers & Operations Research*, 129, 105197. <https://doi.org/10.1016/j.cor.2020.105197>
- Braekers, K., Ramaekers, K., & Van Nieuwenhuysse, I. (2016). The vehicle routing problem: State of the art classification and review. *Computers & Industrial Engineering*, 99, 300–313. <https://doi.org/10.1016/j.cie.2015.12.007>
- Dupas, R., Hsu, T., & Taniguchi, E. (2024). A clustering-routing approach for assigning customers to depots in last mile delivery. *Transportation Research Procedia*, 79, 13–20. <https://doi.org/10.1016/j.trpro.2024.03.004>
- Leung, E. K. H., Das, S., Bektaş, T., & Choi, T.-M. (2026a). Optimizing delivery systems within the e-retail context: a weighted self-organizing map for delivery region partitioning. *European Journal of Operational Research*, 330(1), 100–119. <https://doi.org/10.1016/j.ejor.2025.09.006>
- Leung, E. K. H., Das, S., Bektaş, T., & Choi, T.-M. (2026b). Optimizing delivery systems within the e-retail context: a weighted self-organizing map for delivery region partitioning. *European Journal of Operational Research*, 330(1), 100–119. <https://doi.org/10.1016/j.ejor.2025.09.006>
- Lloyd, S. (1982). Least squares quantization in PCM. *IEEE Transactions on Information Theory*, 28(2), 129–137. <https://doi.org/10.1109/TIT.1982.1056489>

- Maria, E., Budiman, E., Haviluddin, & Taruk, M. (2020). Measure distance locating nearest public facilities using Haversine and Euclidean Methods. *Journal of Physics: Conference Series*, 1450(1), 012080. <https://doi.org/10.1088/1742-6596/1450/1/012080>
- Nie, F., Xue, J., Wu, D., Wang, R., Li, H., & Li, X. (2021). Coordinate Descent Method for k-means. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1–1. <https://doi.org/10.1109/TPAMI.2021.3085739>
- Ou, J., Liu, X., Xing, L., Lv, J., Hu, Y., Zheng, J., Zou, J., & Li, M. (2024). Solving many-objective delivery and pickup vehicle routing problem with time windows with a constrained evolutionary optimization algorithm. *Expert Systems with Applications*, 255, 124712. <https://doi.org/10.1016/j.eswa.2024.124712>
- Ramírez-Villamil, A., Montoya-Torres, J. R., Jaegler, A., & Cuevas-Torres, J. M. (2023). Reconfiguration of last-mile supply chain for parcel delivery using machine learning and routing optimization. *Computers & Industrial Engineering*, 184, 109604. <https://doi.org/10.1016/j.cie.2023.109604>
- Sun, H. J. (Norina), & Arslan, O. (2025). A contextual framework for learning routing experiences in last-mile delivery. *Transportation Research Part B: Methodological*, 194, 103172. <https://doi.org/10.1016/j.trb.2025.103172>
- Vidal, T. (2022). Hybrid genetic search for the CVRP: Open-source implementation and SWAP* neighborhood. *Computers & Operations Research*, 140, 105643. <https://doi.org/10.1016/j.cor.2021.105643>
- Wouda, N. A., Lan, L., & Kool, W. (2024). PyVRP: A High-Performance VRP Solver Package. *INFORMS Journal on Computing*, 36(4), 943–955. <https://doi.org/10.1287/ijoc.2023.0055>