JOURNAL OF DATA SCIENCE
eISSN:2805-5160
Vol.2022:20

Route Optimization based on Clustering and Travelling Salesman Problem

Khadijah Thahira[1], Tri Basuki Kurniawan[1*], Edi Surya Negara[1], Ahmad Haidar Mirza[2], Misinem[3]

[1]Magister of Information Technology, University of Bina Darma, Palembang, Indonesia
[2]Faculty of Computer Science, University of Bina Darma, Palembang, Indonesia
[3]Faculty of Vocational, University of Bina Darma, Palembang, Indonesia

[*]Email: tribasukikurniawan@binadarma.ac.id

Abstract

The rapid pace of e-commerce growth has affected the logistics sectors to face challenges such as the pressure of consumer expectations and increased competition regionally across all players in the supply chain. This e-commerce wave has led the logistics sector to struggle to improve logistics distribution efficiency and reduce operating costs to keep up with consumers' fast-growing demand simultaneously. Therefore, this paper aims to introduce The Route Optimization approaches that are developed to enhance the efficiency of the day-to-day operation in the logistics industry at one leading distribution company, CV. Berkah Express in Palembang, South of Sumatera, Indonesia. The system aims to optimize the last-mile distribution route by reducing the driver's travel distance to drive more efficiently. In this paper, comparison results between the original route taken by the driver and the proposed route based on optimization are conducted and reported. The unsupervised learning result was also achieved and reported in comparison results. Based on the comparison result, the route optimization system was proved effective through example analysis on the said test dataset. The analysis results also reflect how the system's algorithm can provide better routing solutions with shorter distances and lesser time that could decrease the last-mile distribution costs efficiently.

Keywords

Route Optimization, K-Means Clustering, Travelling Salesman Problem, Shortest Path Problem

Introduction

The logistics sector has long heralded the potential of e-commerce. For the past few years, many delivery start-ups have joined the industry. At the same time, more prominent players invested in hubs and partnered with other companies to capitalize on the e-commerce market. According to Global Data's E-Commerce Analytics, the e-commerce wave in Malaysia and Indonesia is

Submission: 10 November 2022
Acceptance: 12 December 2022

estimated to register a 24.7% growth in 2020 (Sitorus et al., 2021). Surprisingly, it is expected to reach US$12.6 billion by 2024, increasing at a compound annual growth rate (CAGR) of 14.3% between 2020 and 2024. Such brisk expansion would mean a tremendous market for the logistics sector. Still, it would also mean the pressure of consumer expectation, increased competition, and lack of customer loyalty. With the logistics sector being the key differentiating factor between e-commerce companies, organizations must focus on strengthening their last-mile delivery services. Opportunist logistics players are started to rapidly embark on their digitalization journey to keep up with the fast-growing demand prompted by the explosive growth in e-commerce (Pham et al., 2019).

The CV. Berkah Express, one of Palembang's leading technology-empowered logistics platforms, has proposed a solution to ensure seamless end-to-end data flow from the hub to the consumers known as The Route Optimization Project. The inspiration to develop in-house Route Optimization came from the Travelling Salesman Problem (TSP), Vehicle Routing Problem (VRP), and Cargo Load Problem (CLP) faced by Berkah Express in its day-to-day operation. It is an initiative to solve the last-mile distribution problem by shortening the delivery time to drive greater efficiency to compete regionally across all players in the supply chain. Technically, route optimization refers to a solution for so-called vehicle routing problems (VRPs). The Vehicle Routing Problem or VRP is the challenge of designing optimal routes from a depot to a set of destinations, each with business-specific constraints, such as time windows, vehicle limitations, cost controls, and others (Marampoutis et al., 2022).

The time required to solve VRPs is growing relative to the size of the problem. For significant issues, it could take years to find the optimal solution. These are the complex problems that operations teams face, making it even impossible to plan the optimal route solution quickly and manually. It took more than 1 hour for The Berkah Express Ops Planning Team to devise the optimal route for 100 drop-off points. As the order numbers increased, the planning time grew exponentially. Other than higher operating costs, inefficient route and load planning also contribute to traffic congestion and an increase in carbon footprint (Laporte, 2007). Therefore, by developing the route optimization project, it would be easier to automate and generate an effective optimal route plan to reduce the operation workloads and to operate distribution costs efficiently (Sahoo et al., 2005).

The routing of vehicles and scheduling of deliveries are crucial for supply chain operations, as both determine to a great extent, the distribution costs and customer satisfaction. Because e-commerce is growing, and the demand for distribution goods is increasing, the vehicle routing problem has become one of the most studied topics in operational research (Konstantakopoulos et al., 2022). Researchers have studied to identify the different variants of the VRPs and proposed algorithms to solve the problems to achieve good efficiency and optimal distribution costs.

Capacitated Vehicle Routing Problem (CVRP) is considered the simplest variant of the VRP as the capacity of the vehicles is the only constraint considered, which is assumed to be identical in terms of actual load capacity and costs. Altabeeb et al. (2019) have proposed a new hybrid Firefly Algorithm (FA) called CVRP-FA to solve the capacitated vehicle routing problem. The researchers mentioned that in CVRP-FA, FA is integrated with two types of local search and

Submission: 10 November 2022
Acceptance: 12 December 2022

genetic operators to enhance the solution's quality and good in accelerating the convergence. The results presented that CVRP-FA has a fast convergence rate and high computational accuracy. Based on the development, the algorithm can be applied to solve other VRPs, such as VRP with time windows or different combinatorial optimization problems.

According to a journal post by Liu (2020), the researcher worked on a route optimization model for rural e-commerce logistics' (RECL) last-mile distribution to maximize the logistics company's profit. At the same time, they kept RECL's features in mind, such as its extended supply chain and low consumption density. Liu (2020) and Li et al. (2019) proposed Ant Colony Optimization (ACO) algorithm was improved to suit the physical distribution of goods by applying an innovative approach to updating the pheromone that results in a better solution to solve the VRP. Both researchers demonstrated satisfying performance with the improved ACO (IACO), which has better solution quality than conventional ACO.

To increase the reliability of the combinatorial optimization algorithm for solving the VRPs, the address matching and the geocoding process is also crucial to providing accurate distribution of consumers' geographic coordinates. Lee et al. (2020) mentioned that to enhance address matching on street-based addresses implementing the address geocoding algorithm using machine learning. The developed address geocoding algorithm contains three main modules: address parsing, address matching, and address locating. The researchers stated that address matching using machine learning could cope with human errors, such as spelling and input addresses, to match existing addresses accurately.

This approach of building structured geocoding algorithm modules can be applied to any other geocoding system to precisely convert addresses into geospatial data for relevant applications such as delivery routing and consumer distributions.

Address parsing has always been challenging as the address structure is complex and sometimes composed of anomalies that lead to inaccurate results. Parsing is the analysis of sequence characters and breaks that given text into meaningful pieces (Lee et al., 2020). This process of parsing is highly related to the task called information extraction. It is a process of extracting structured information from unstructured text, and this task can quickly be done by using Named Entity Recognition (NER).

Nasar et al. (2021) defined NER as an information extraction task concerned with finding textual mentions of entities belonging to a predefined set of categories. Tarcar et al. (2020) proved that NER could extract phrases related to pharmaceutical chemicals with dosage, disease, and symptoms from Electronic Health Records (EHR). Parsing addresses with NER will be a practical approach to extract the entities seamlessly if the availability of reasonable-sized and high-quality annotated datasets is provided.

Methodology

Submission: 10 November 2022
Acceptance: 12 December 2022

The actual data from CV. Berkah Express has already been gathered and prepared, which includes a week's worth of transactional data for the customer's item that should be delivered, as shown in Table 1.

Table 1. Experiment data

Day #	Date	No of drivers	Total of Items
1	2022-01-24	7	154
2	2022-01-25	5	159
3	2022-01-26	5	143
4	2022-01-27	9	221
5	2022-01-28	8	192
6	2022-01-29	5	111
7	2022-01-30	6	93
	Total	45	1073

Experiment Setup

The experiment began by calculating the distance each driver took to deliver the customer items and compared it with the distance the optimization result used Google RO Tools proposed for that driver.

The second experiment compares the total original driver travel distance for each day with the total travel distance after the clustering was applied, as shown in Figure 1.

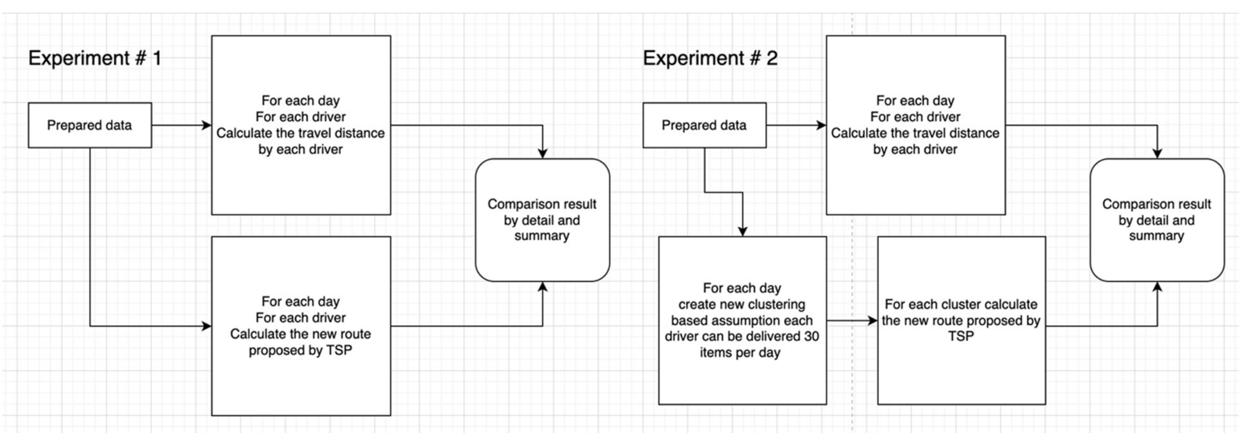


Figure 1. Experiment setup

Figure 1 shows the experiment was conducted, and the comparison result was reported. The conclusion will be based on two (2) experiments and comparison results. The investigation was conducted using Python language at The Jupiter notebook on workstations with Intel Xeon Processor 3.3 GHz and 4 GB RAM.

Submission: 10 November 2022
Acceptance: 12 December 2022

Google OR Tools

OR-Tools is an open-source optimization software package designed to tackle the world's most challenging issues in vehicle routing, flows, integer and linear programming, and constraint programming (Google, 2022).

After modeling the issue in your preferred programming language, you may solve it with one of a half-dozen solvers: commercial solvers like Gurobi or CPLEX or open-source solvers like SCIP, GLPK, or Google's GLOP and award-winning CP-SAT.

The initial version was uploaded on GitHub in September 2015, and version 7.2 was released in July 2019. While the Google AI-based solution was developed in C++, the suite may also be utilized with Python, Java, and C#. Linux, Mac, and Windows versions are all accessible for download and installation.

Unsupervised Learning

The clustering algorithm in this project is proposed to keep the cluster member items close to each other before the optimization route is proposed. In many cases, we need to be ensured that our cluster result should have an equal number of member clusters. Clustering with the confined issue can be utilized to accomplish this. One of the python modules that can be used is K-Means constrained from the PyPI project (Levy-Kramer, 2022). In this scenario, we need each driver to deliver 30 items and parcels per trip.

Results and Discussion

The first comparison that will be discussed is the total distance of each driver per day based on the original route by a driver with the distance route by proposed optimization algorithms. Table 2 and Figure 2 display the grouping results and comparison between the two results.

Table 2. Comparison results for experiment 1

Day #	Date	No of drivers	Total of Items				
1	2022-01-24	7	154				
	No	Driver Name	# items	By driver (km)	By TSP (km)	Different (km)	(%)
	1	Driver 01	13	25.19	21.83	3.36	-13.3
	2	Driver 02	18	42.03	33.06	8.96	-21.33
	3	Driver 03	12	17.96	13.31	4.66	-25.92
	4	Driver 04	53	106.38	75.39	31.00	-29.14
	5	Driver 05	19	13.78	14.98	-1.20	+8.71
	6	Driver 06	25	87.50	39.47	48.02	-54.89
	7	Driver 07	14	77.55	78.12	-0.57	+0.74
	Total		154	370.40	276.17	94.23	-25.44

Submission: 10 November 2022
Acceptance: 12 December 2022

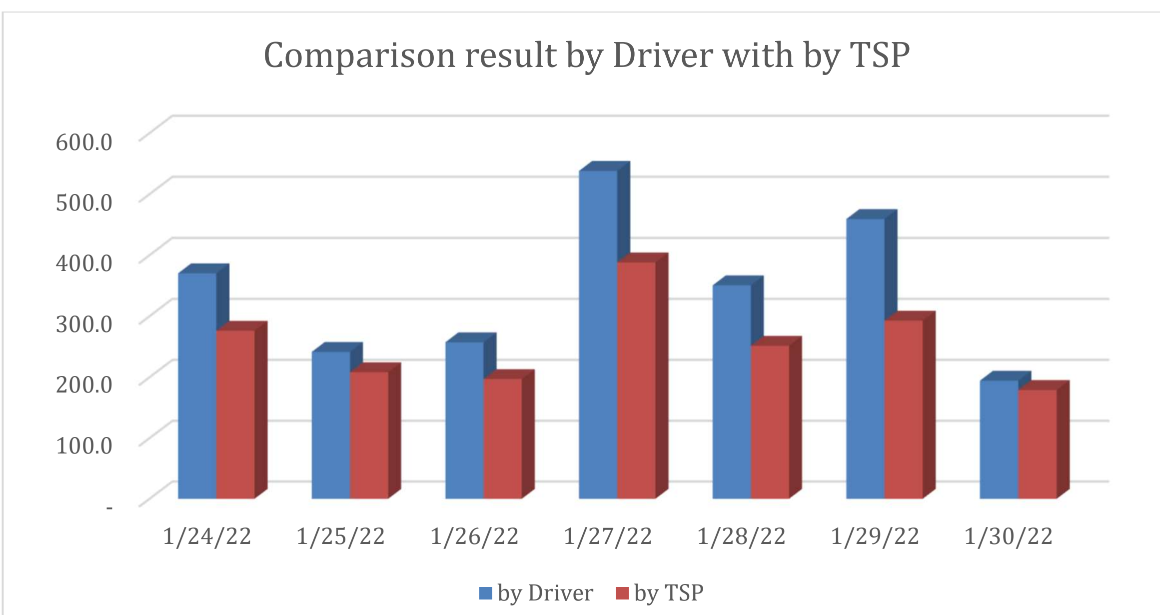
JOURNAL OF DATA SCIENCE
eISSN:2805-5160
Vol.2022:20

2	2022-01-25	5	159				
	No	Driver Name	# items	By driver (km)	By TSP (km)	Different (km)	(%)
	1	Driver 01	33	67.08	39.77	27.31	-40.71
	2	Driver 02	62	70.26	51.76	18.50	-26.33
	3	Driver 03	37	57.42	60.26	-2.84	+4.94
	4	Driver 04	20	32.60	45.29	-12.68	+38.90
	5	Driver 05	7	14.22	11.10	3.12	-21.91
	Total		159	241.59	208.19	33.40	-13.83
3	2022-01-26	5	143				
	No	Driver Name	# items	By driver (km)	By TSP (km)	Different (km)	(%)
	1	Driver 01	45	68.09	52.44	15.64	-22.98
	2	Driver 02	8	5.85	5.87	-0.02	+0.40
	3	Driver 03	18	37.92	31.32	6.06	-17.40
	4	Driver 04	29	50.99	44.60	6.39	-12.53
	5	Driver 05	43	92.52	62.63	31.89	-33.53
	Total		143	257.36	196.87	60.49	-23.50
4	2022-01-27	9	221				
	No	Driver Name	# items	By driver (km)	By TSP (km)	Different (km)	(%)
	1	Driver 01	24	24.77	17.20	7.57	-30.56
	2	Driver 02	18	37.45	39.92	-2.47	+6.61
	3	Driver 03	22	83.80	53.27	30.53	-36.43
	4	Driver 04	24	40.46	32.26	8.20	-20.26
	5	Driver 05	63	152.08	92.29	59.79	-39.31
	6	Driver 06	18	58.82	38.94	19.88	-33.80
	7	Driver 07	19	38.94	31.27	7.67	-19.71
	8	Driver 08	14	73.42	67.18	6.25	-8.51
	9	Driver 09	19	28.33	15.82	12.51	-44.16
	Total		221	538.06	388.14	149.92	-27.86
5	2022-01-28	8	192				
	No	Driver Name	# items	By driver (km)	By TSP (km)	Different (km)	(%)
	1	Driver 01	25	22.71	13.56	9.15	-40.29
	2	Driver 02	41	64.19	33.60	30.59	-47.65
	3	Driver 03	12	34.02	32.57	1.45	-4.26
	4	Driver 04	34	53.35	29.29	24.06	-45.09
	5	Driver 05	12	13.82	13.73	0.09	-0.66
	6	Driver 06	12	23.30	18.07	5.24	-22.47

Submission: 10 November 2022
Acceptance: 12 December 2022

JOURNAL OF DATA SCIENCE
eISSN:2805-5160
Vol.2022:20

	7	Driver 07	29	80.07	61.01	19.06	-23.80
	8	Driver 08	27	59.12	49.71	9.41	-15.91
	Total		192	350.58	251.55	99.03	-28.25
6	2022-01-29	5	111				
	No	Driver Name	# items	By driver (km)	By TSP (km)	Different (km)	(%)
	1	Driver 01	38	178.15	54.26	123.89	-69.54
	2	Driver 02	12	12.31	10.92	1.39	-11.30
	3	Driver 03	13	59.03	24.86	34.17	-57.89
	4	Driver 04	20	98.15	143.75	-45.59	+46.45
	5	Driver 05	28	111.54	58.89	52.64	-47.20
	Total		111	459.18	292.68	166.50	-36.26
7	2022-01-30	6	93				
	No	Driver Name	# items	By driver (km)	By TSP (km)	Different (km)	(%)
	1	Driver 01	13	23.90	18.50	5.40	-22.59
	2	Driver 02	20	27.68	24.06	3.63	-13.10
	3	Driver 03	13	32.63	27.90	4.72	-14.48
	4	Driver 04	18	41.47	44.44	-2.97	+7.16
	5	Driver 05	13	41.68	39.53	2.14	-5.14
	6	Driver 06	16	27.08	24.79	2.29	-8.46
	Total		93	194.44	179.23	15.21	-7.82
Total			1073	2411.61	1792.83	618.79	-25.66



Submission: 10 November 2022
Acceptance: 12 December 2022

Figure 2. Comparison results for experiment 1 total distance by Driver with by TSP

Table 2 shows a few drivers get the better route (ex., Driver 05 and driver 06 on January 24, 2022) compared with the proposed by TSP, but generally, the Optimization result obtained a better route. The total distance for each date shows the TSP offered the better route. For all 7 (seven) days, the travel length can be reduced by about 25.66 %. But the higher performance shows on January 29, 2022, 36.26 %, and the lower performance shows on January 30, 2022, about 7.82%. Figure 2 shows the comparison of the total distance obtained by Driver and by TSP. Overall, the travel distance obtained by TSP for each date gets the better route.

The second comparison was conducted between the original route made by the driver with the result of new clustering generated using K-Means Constraint algorithms based on the assumption each driver can be delivered a maximum of 30 items per trip. For the new proposed route, first, we create a new cluster which each cluster as a new driver. Each driver will have a few members of an item the driver will deliver, as shown in Table 3 and Figure 3.

Table 3. Comparison results for experiment 2

Day #	Date	No of Driver	By Driver (km)	By TSP (km)	Total of Items	No of Cluster		
1	24/01/2022	7	370.40	276.17	154	5		
						Cluster #	# items	By Clustering +TSP (km)
						1	27	34.01
						2	33	32.28
						3	33	82.34
						4	33	49.61
						5	28	43.95
							Total	242.19
2	25/01/2022	5	241.59	208.19	159	5		
						Cluster #	# items	By Clustering +TSP (km)
						1	33	71.39
						2	31	43.34
						3	33	39.77
						4	33	61.98
						5	29	42.23
							Total	258.72
3	26/01/2022	5	257.36	196.87	143	5		
						Cluster #	# items	By Clustering +TSP (km)
						1	33	42.66
						2	27	49.85
						3	27	37.98

Submission: 10 November 2022
Acceptance: 12 December 2022

JOURNAL OF DATA SCIENCE
eISSN:2805-5160
Vol.2022:20

							4	27	36.79
							5	29	34.30
								Total	201.59
4	27/01/2022	9	538.06	388.14	221	7			
							Cluster #	# items	By Clustering +TSP (km)
							1	33	23.22
							2	33	23.01
							3	27	46.74
							4	33	25.64
							5	29	51.22
							6	33	89.77
							7	33	56.38
								Total	315.98
5	28/01/2022	8	350.58	251.55	192	6			
							Cluster #	# items	By Clustering +TSP (km)
							1	33	20.16
							2	32	44.75
							3	28	49.71
							4	33	31.14
							5	33	49.55
							6	33	16.77
								Total	212.08
6	29/01/2022	5	459.18	292.68	111	4			
							Cluster #	# items	By Clustering +TSP (km)
							1	27	48.83
							2	27	178.91
							3	27	40.62
							4	30	50.83
								Total	319.19
7	30/01/2022	6	194.44	179.23	93	3			
							Cluster #	# items	By Clustering +TSP (km)
							1	29	52.05
							2	33	45.40
							3	31	39.15
								Total	136.61

Submission: 10 November 2022
Acceptance: 12 December 2022

Total	45	2411.61	35	1686.35
Different			10	725.26
%			-22.22	-30.07

Table 3 shows one day, the original driver got the better route (for data date 25/01/2022) compared with the proposed by Clustering +TSP, but generally, the Optimization result obtained a better route. The total distance got by the original driver is 2411.61 km compared with the distance from Clustering + TSP at 1686.35. A different distance is 725.26 km means to get a decrease of 30.07%. The optimization result got a smaller number of drivers to compare with the original result by driver. The original result by driver needs 45 drivers for 7 (seven) days and the optimization result by clustering + TSP only needs 35 clustering or drivers, which means a reduction of 22.22%.

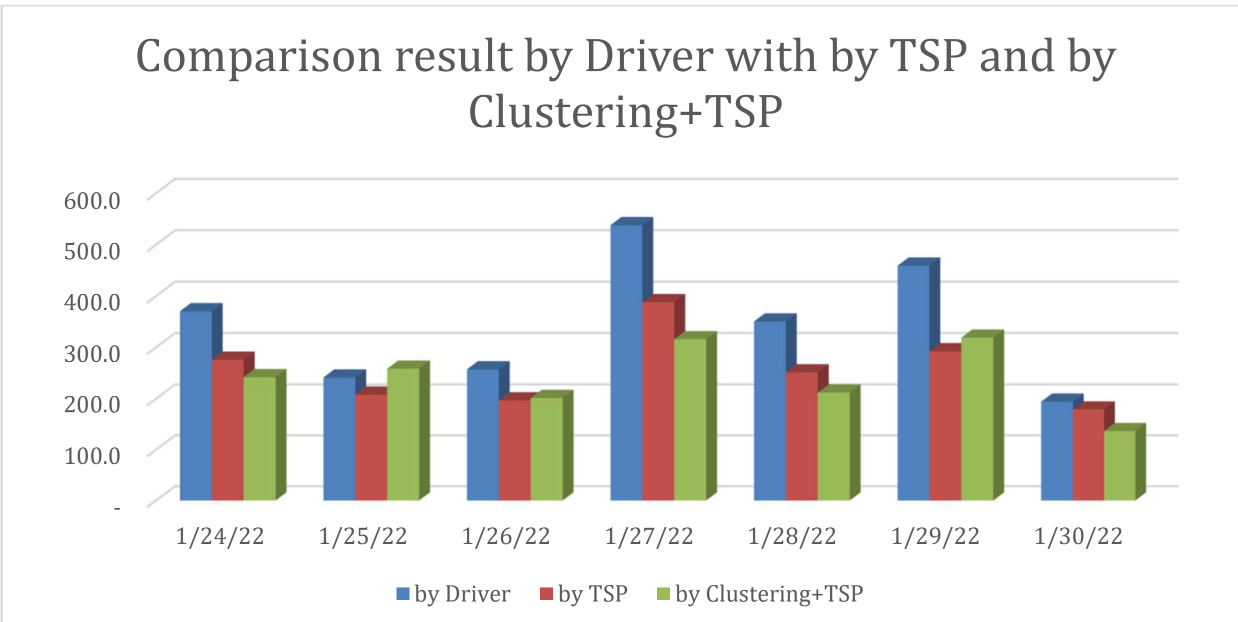


Figure 3. Comparison results for experiment 2 total distance by Driver with by TSP and by Clustering + TSP

Figure 3 shows the comparison of the total distance obtained by the Driver with the result by TSP and by Clustering + TSP. Overall, the best result is the travel distance obtained by Clustering + TSP for all 7 (seven) days.

The final result of the experiment is a visualization using a map to show the clustering has been created are close to each other, as shown in Figure 4.

Submission: 10 November 2022
Acceptance: 12 December 2022

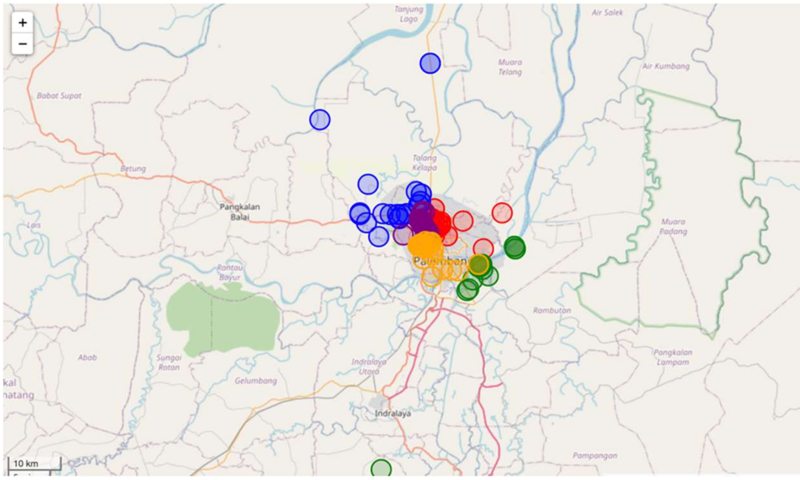


Figure 4. Visualization result to show the position of each clustering member's data

Conclusion

This study intends to introduce The Route Optimization methodologies established to improve the effectiveness of the ongoing operations in the logistics sector. By minimizing the driver's trip distance so that they can drive more effectively, the system seeks to optimize the last-mile distribution route. In this study, comparisons between the driver's original route and the suggested path based on optimization are made and the findings are presented. The route optimization method was shown to be effective by example analysis on the mentioned test dataset based on the comparison result. The analysis's findings also demonstrate how the system's algorithm may offer superior routing solutions with shorter distances and less time, which might effectively lower the costs associated with last-mile distribution.

References

Altabeeb, A. M., Mohsen, A. M., & Ghallab, A. (2019). An improved hybrid firefly algorithm for capacitated vehicle routing problem. *Applied Soft Computing*, *84*, 105728. https://doi.org/https://doi.org/10.1016/j.asoc.2019.105728

Google. (2022). Google OR Tools. Retrieved October 23, 2022, from google.com website: https://developers.google.com/optimization

Konstantakopoulos, G. D., Gayialis, S., & Kechagias, E. (2022). Vehicle routing problem and related algorithms for logistics distribution: a literature review and classification. *Operational Research*, 1–30. https://doi.org/10.1007/s12351-020-00600-7

Laporte, G. (2007). What You Should Know About the Vehicle Routing Problem. *Naval Research Logistics (NRL)*, *54*, 811–819. https://doi.org/10.1002/nav.20261

Lee, K., Claridades, A. R., & Lee, J. (2020). Improving a Street-Based Geocoding Algorithm Using Machine Learning Techniques. *Applied Sciences*, *10*, 5628. https://doi.org/10.3390/app10165628

Levy-Kramer, J. (2022). k-means-constrained 0.7.2. Retrieved from https://pypi.org/project/k-means-constrained/

Li, Y., Soleimani, H., & Zohal, M. (2019). An improved ant colony optimization algorithm for the multi-depot green vehicle routing problem with multiple objectives. *Journal of Cleaner Production*, *227*, 1161–1172. https://doi.org/https://doi.org/10.1016/j.jclepro.2019.03.185

Liu, W. (2020). Route Optimization for Last-Mile Distribution of Rural E-Commerce Logistics Based on Ant Colony Optimization. *IEEE Access*, *PP*, 1. https://doi.org/10.1109/ACCESS.2020.2964328

Marampoutis, I., Vinot, M., & Trilling, L. (2022). Multi-objective vehicle routing problem with flexible scheduling for the collection of refillable glass bottles: A case study. *EURO Journal on Decision Processes*, *10*, 100011. https://doi.org/https://doi.org/10.1016/j.ejdp.2021.100011

Nasar, Z., Jaffry, S. W., & Malik, M. (2021). Named Entity Recognition and Relation Extraction: State of the Art. *ACM Computing Surveys*, *54*. https://doi.org/10.1145/3445965

Pham, H. C., Nguyen, D., Doan, C., Thai, Q., & Nguyen, N. (2019). Last Mile Delivery As a Competitive Logistics Service-A Case Study. *Proceedings of the 9th International Conference on Operations and Supply Chain Management (OSCM)*, (Lmd), 1–8. Retrieved from https://researchrepository.rmit.edu.au/esploro/outputs/conferenceProceeding/Last-Mile-Delivery-as-a-Competitive-Logistics-Service--A-Case-Study/9921861060001341

Sahoo, S., Kim, S., Kim, B.-I., Kraas, B., & Jr, A. (2005). Routing Optimization for Waste Management. *Interfaces*, *35*, 24–36. https://doi.org/10.1287/inte.1040.0109

Sitorus, A., Gultom, M., & Andini, V. (2021). E-commerce boosts the logistics sector in Indonesia - Opinion - The Jakarta Post. Retrieved October 21, 2022, from Thejakartapost.Com website: https://www.thejakartapost.com/academia/2021/01/03/e-commerce-boosts-the-logistics-sector-in-indonesia.html

Tarcar, A. K., Tiwari, A., Rao, D., Dhaimodker, V. N., Rebelo, P., & Desai, R. (2020). Healthcare NER models using language model pretraining. *CEUR Workshop Proceedings*, *2551*(Hsdm), 12–18. https://doi.org/10.1145/3336191.3371879

Submission: 10 November 2022
Acceptance: 12 December 2022