

# Swarm

*by* Silvia Rostianingsih

---

**Submission date:** 25-Jul-2024 09:19AM (UTC+0700)

**Submission ID:** 2422062168

**File name:** IJIRSET\_SILvia.pdf (1.03M)

**Word count:** 4129

**Character count:** 21260

# Particle Swarm Optimization Algorithm for Vehicle Routing Optimization

Silvia Rostianingsih, Andreas Handojo, Jason

Associate Professor, Informatics Department, Petra Christian University, Surabaya, Indonesia

Associate Professor, Informatics Department, Petra Christian University, Surabaya, Indonesia

Student, Informatics Department, Petra Christian University, Surabaya, Indonesia

**ABSTRACT:** Technology has been one of the critical factors behind the industrial revolution. Companies must now use technological assistance and data processing to produce faster and more efficient business processes. Our case study is using HDPE Plastic Company, which is in Surabaya, is trying to handle the increasing frequency of shipments. Due to the rising frequency of shipments, the company is often overwhelmed in handling its loads because no system can quickly determine the shipping route.

Moreover, other route-determining factors such as shipment weight, truck capacity, and unique delivery hour requests, manually add to the route's complexity. The system will run the K-Means cluster function from the database to cluster all customers in the company. This cluster is one of the factors determining the fitness value in the Particle Swarm Optimization (PSO) algorithm. After the order data is obtained, the system will use the PSO algorithm to determine the delivery agenda for each truck. The determining factors of PSO include customer location, priority hours of customer requests, order weight, and loading capacity of different types of trucks. After obtaining the delivery table of each truck, the system will use the help of Google Directions Service to determine the routing order from each truck.

The result of this system is a delivery route optimization system that can provide route selection recommendations for each truck in the company. The system is also able to sort shipments with various shipping priority restrictions. From the test results, the PSO algorithm in the system can produce routes with less total distance traveled and less travel duration than the routes generated manually by the employees in the company.

**KEYWORDS:** Google Directions Service; Particle Swarm Optimization; Vehicle Routing Problem.

## I. INTRODUCTION

HDPE Plastic Company has around 60 product variations, each with different sizes, colors, and weights. HDPE Plastic Company implements a make-to-stock strategy to meet customer demand, requiring a large production capacity and extensive inventory. In a day, the company can produce 1300 bales with the help of 26 extruder machines and a workforce of around 600 people.

The company uses ten trucks of different sizes for the distribution of the product. Orders in the East Java region is sent directly to the destination with a fleet of trucks. Whereas orders in the East Java region will be delivered by the company using its fleet to the customer's selected courier, and orders from outside East Java to outside the island will be delivered by the company's trucks to container depots to be loaded to containers and transported by ships.

This research will create a route optimization system using the Particle Swarm Optimization (PSO) algorithm to determine the most optimal list of delivery routes for each truck in the company. The route for the company is a single route where shipments only depart from the company's warehouse once a day. Google Maps API is used to calculate the distance and time from each destination and provide a visualization of the route given to each truck to make it easier for the driver. The result is to help companies determine the optimal distance to provide the most efficient delivery time and reduce overtime pay and fuel costs. This system will be accessible to the admin in the company office and the truck driver to update the delivery status when it arrives at a specific destination. This system will also provide reporting regarding delivery performance within the company. Tests on this system will be carried out by comparing the differences in the results of distance, time, and costs required if the route is determined manually with the system that will be created using the PSO algorithm in this research.

## II. RELATED WORK

20 PSO can be used for solving a routing problem in the urban multi-modal network using route length, traffic, comfort, and safety as objective function [1]. Various previous studies discuss route problems and how to optimize them. Study by Hannan presents the CVRP problem for the route for picking up waste from the waste collection site, which is later sent to the waste processing site [2]. The proposed method uses the PSO algorithm, but this study's drawbacks are that the route's determinants are only distance and are not equipped with time windows. Another study by Mussagulova raised the VRP problem using the PSO algorithm. However, 10 the model analyzed in this study differs from this research, where the research focuses on the model problem of the Vehicle Routing Problem with Simultaneous Pickup and Delivery (VRPSD) and Capacitated Vehicle Routing Problem with Time Windows (CVRP-TW) [3]. There is also a study that uses improved Genetic Algorithm (GA) for the route optimization, 12 system that applied to Vehicle Routing Problem Pick-up and Delivery with Time Windows (VRPPDTW) [4] and Vehicle Routing Problem with Cross-Docking and Carbon Emissions Reduction [5]. The improved GA comes with a shorter computation time than standard GA. Other research is using Sweep Nearest method to cluster data based on nearest neighbor concept [6] [7] [8]. Transportation problems also can be solved by optimization heuristics and meta-heuristic algorithms [9].

## III. METHODOLOGY

### 1. Particle Swarm Optimization (PSO)

11 Particle Swarm Optimization (PSO), first discovered by Kennedy and Eberhart in 1995, is a computational method that optimizes a problem by having a population called particles. 18 These particles are moved in a certain space and at random locations. All particles 5 have velocity and position [10]. The Particle Swarm Optimization algorithm simulates the process of foraging by birds and considers the solution space of the problem as the space for feeding birds. In PSO, the solution to the optimization problem is the position of the particle in the search space. PSO has several advantages, such as relatively few parameters, 6 to tune, low computational time, higher efficiency, and probability to find the global optima [11]. This makes the PSO algorithm suitable for solving problems that have many combinations of factors, such as shipping problems, manufacturing problems, and scheduling problems [12].

This PSO starts with the presence of a population (Particle) that has a random location. All these particles will be given a random speed to reach the most suitable location (14). Each location will be evaluated to determine whether the particle's position is the best based on the objective. The best fitness value for each particle will be stored in the Personal Best (pBest). Every time these particles move, they will be evaluated and determined which has the 6 st fitness value and stored in the Global Best (gBest).

The new velocity of the particle will be shown by equation 1, while the new location of this particle can be shown by equation 2.

$$v_i = W v_i + c1r1 (Pbest_i - x_i) + c2r2 (Gbest - x_i) \quad (1)$$

$$x_i = x_i + v_i \quad (2)$$

3 However, segmentation PSO algorithm can be used to divide the initial values into segments to help PSO particles during the search for the optimal values finding a local best position that may the global best position is around it [13].

### 2. PSO for the company

The system flow will begin with the input process for all customer orders that day. Order details include the address for delivery, customer name, the number of items purchased, method of payment, and request for special delivery hours, if any. 7 If there is new customer data, the system will add new customer data and re-cluster the entire customer database. This cluster is used as one of the determining factors for fitness value in the PSO algorithm. After all orders on that day have been recorded, the system will proceed to the following process. In this process, the system will use the Particle Swarm Optimization algorithm to get each truck's agenda or delivery table. After the delivery table is determined, the system will use the help of the Google Directions Service to get the route order from the delivery table for each truck generated by the PSO algorithm in the previous process. After the sequence of routes is obtained, the system will display route destinations for the drivers. Visualization page will be obtained with the help of the Google Maps API. The main algorithm can be seen in Fig 1.

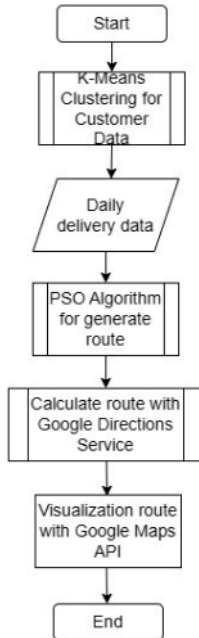


Fig 1. Main algorithm

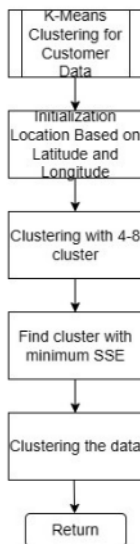


Fig 2. K-Means clustering algorithm

Fig 2 is the process flowchart of K-Means Clustering. The system will receive input from customer location data, including the customer's latitude and longitude. Then the system will carry out the K-Means Clustering process on

the customer's longitude and latitude data. It will initially group the customer data using a different number of clusters, ranging from 4 clusters to 8 clusters. After that, it will check each cluster's Error of the Sum Squares (SSE) value. The number of clusters with the smallest SSE value will be selected, and customer data will be grouped according to the number of clusters. The clustering function will be generated whenever additional data is in the customer database. Data from this customer cluster will be used for the determinants of the PSO algorithm. K-Means is utilized for unsupervised learning. While for data that can be classed such as sentiment analysis, it can be done with supervised learning such as Support Vector Machine [14].

Fig 3 is the PSO algorithm. The way the PSO algorithm works itself is to start by setting the particle object itself. This particle will have content in the form of delivery and velocity, which is as much as the number of existing shipments. The algorithm's initialization is to determine each shipment on each truck randomly. Velocity in the algorithm is also determined randomly or randomly with a limit on the number of trucks. This velocity is later useful in determining the movement of the contents of this particle in each iteration; the movement in question is the combination of each shipment getting which truck ID. The movement of this velocity is also influenced by other variables, namely inertia weight and two constant variables  $c1$  and  $c2$ . For this system, the inertia weight is set at 0.729, and the constant numbers  $c1$  and  $c2$  are 2.05. This movement is executed using formula one and formula 2. In this algorithm, if the  $X_i$  results obtained are more than the total number of trucks, then the number will be modulated before being used for  $X_i$  in the next iteration.

After setting the particle, it will be executed according to the number of iterations. However, the system also sets a stop criterion where if there is no combination with a better fitness value in the 500 iterations since the best combination was found, the system will stop the iteration.

At each iteration, the fitness value of the current combination will be sought, and if it is better than the existing one, the combination will be stored. The initial fitness value of the particle is determined by the number 9999999. In assessing the fitness value of this algorithm, the resulting fitness value is the smaller, the better. For the assessment of the particle fitness value of this algorithm, it is determined by several factors, including:

- Fitness starts with the number 0.
- The total duration of sending, the fitness value will be added to the total duration of sending (in seconds) in that combination.
- The total distance sent, the fitness value will be added to the total distance sent (in meters) for that combination, divided by 10.
- For each truck that exceeds the duration limit on the delivery trip, the fitness will be increased by 1000000. The duration on the road is set at (3 hours or 10800 seconds).
- Number of priority hours: If trucks have more priority hours than the average, the excess number will be multiplied by 5000.
- For each truck, if one exceeds the load, 200000 will be added.
- For each truck, if it does not exceed the load, the amount of the difference between the load and the capacity of the truck will be added.
- For each truck, if one delivers to more than 1 cluster, 10,000 will be added to the number of clusters that must be sent.

Some research is used other other swarm algorithms such as Binary Grey Wolf Optimizer (GWO), which has proved that more effective and efficient for bigger dataset [15].

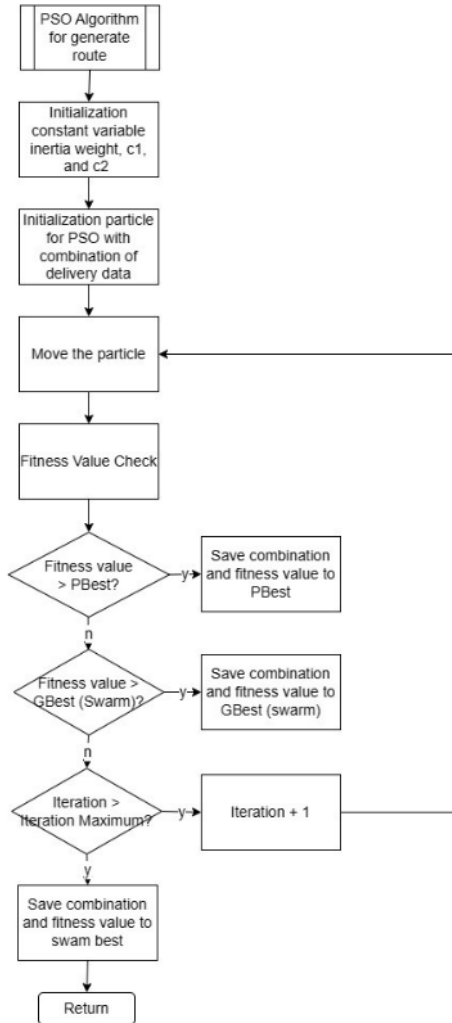


Fig 3. PSO algorithm

After executing the PSO algorithm, an agenda or shipment table will be generated for each truck. However, from the list of destinations in each shipment table, the order of routes for each destination has yet to be determined. The order of this route will be determined using the help of the Google Maps Waypoints API and an algorithm to meet the current time windows or priority hours. Fig 4 is a process flowchart image of the process of determining the route order for each of these trucks.

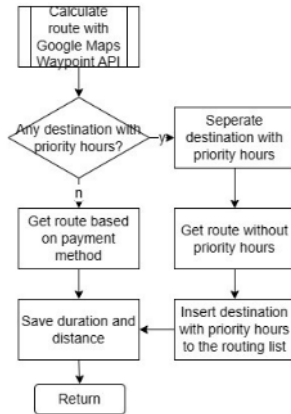


Fig 4. Routing algorithm

### 3. Google Directions Service

The DirectionsService object is used to communicate with the Google Maps API Directions Service which accepts requests for direction and returns an efficient path. Travel time is the main factor that is optimized, but other factors such as distance may be considered. The service also can make waypoints along the route, which is a mark for beginning, end, and intermediate stops.

## IV. EXPERIMENTAL RESULTS

Our experiment performed the K-Means clustering function to group customers based on their location. This clustering function is performed by grouping latitude and longitude attributes of the customers. The number of clusters assigned to this function is six, covering East Surabaya, Central Surabaya, West Surabaya, South Surabaya, North Surabaya, and outside Surabaya. Through this function, the system manages to group customers as desired, and the results of this clustering will be displayed on.

### 1. POS algorithm speed testing

The data in the table shows that shipments with only ten destinations require only about two seconds to produce the results of the delivery table for each truck. In contrast, shipments with 50 destinations require up to seven times longer, around 14.5 seconds. The number of priority hours also determines the speed of the calculation process of this algorithm, where the more priority hours there are, the longer it will take. However, priority hours only matter slightly compared to the number of destinations. Like the total running time of the algorithm, the time to determine the fitness value for one combination or particle also increases when the number of destinations increases.

Fig 5. is a comparison graph between the number of destinations and the total response time generated by the algorithm. From the graph, the increase is almost linear or resembles a straight line. There will be a linear increase of about 3 seconds for every ten new destinations.

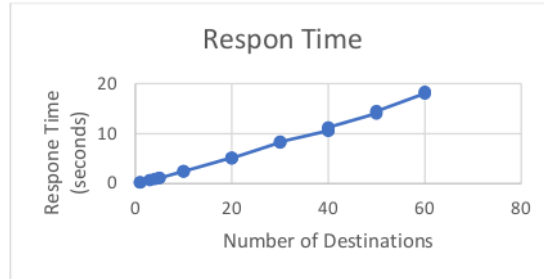


Fig 5. Respon time

2. Route calculation without time windows

This test will be carried out to find out whether the system can run if there is no priority hour at all on the agenda or delivery table for one truck. The shipment data in this test are in Table 1.

In Table 1, there is complete data regarding the destinations that the trucks must go to in a day. Data includes ID, full name, address, latitude, longitude, and payment method. ID is used to identify the shipment on the system. Latitude, longitude, and payment method data will be sent to the Google Waypoints API to run the runGoogleWaypoints. The sequence of the route that has been optimized can be seen in Table 2 and Fig 6.

Table 1. Data destination for one truck

ID	Area	Latitude	Longitude
420	Krembangan	-7.22169405583	112.723391065
422	Pabean Cantikan	-7.21963194151	112.736409451
425	Kembang Jepun	-7.2374599360	112.74034843
432	Gubeng	-7.28920546385	112.7614879

Table 2. The route without time windows

Priority	Area	Time Departure	Time Arrival	Duration (minutes)	Distance (km)
1	Pabean Cantikan	08:00:00	08:22:53	23	9.7
2	Krembangan	08:52:53	09:00:52	8	2.8
3	Kembang Jepun	09:30:52	09:43:33	13	5.7
4	Gubeng	10:13:33	10:34:39	22	8

3. Route calculation with time windows

This test will be carried out to find out whether the system can run with priority hour for one truck. The shipment data in this test is in Table 3. The system will run as without prior priority hours. Then the system checks whether there is one of the sequences of routes that can be inserted with shipments that have priority hours when viewed from the time of departure. It will check whether time windows correspond to the priority hours. If there are, destinations with priority hours will be tucked in. The system will be restarted to recalculate the delivery duration and distance from each destination. The algorithm can be seen in Fig. 7. The results of the first test regarding shipments with priority hours are shown in Table 4.



Route 1:  
 Distance 9.7 km  
 Duration: 23 minutes  
 Departure: Kalikepiting No. 175, Kel Pacar Kembang, Tambaksari  
 Arrival: Kalimas Baru I No 25, Perak Utara, Kec Pabean Cantikan

Route 2  
 Distance 2.8 km  
 Duration 8 minutes  
 Departure: Kalimas Baru I No 25, Perak Utara, Kec Pabean Cantikan  
 Arrival: Tanjung Batu No 4, Perak Barat, Kec Krembangan

Fig 6. Visualization of the route without time windows

Table 3. Data destination for one truck

ID	Area	Latitude	Longitude	Priority Hours
420	Pengampon	-7.246367760	112.747356	N/A
422	Margomulyo	-7.253728491	112.6933163	N/A
425	Kalimati	-7.236719627	112.740030	N/A
432	Gembong Sawah	-7.24092498	112.7500436	09:00:00

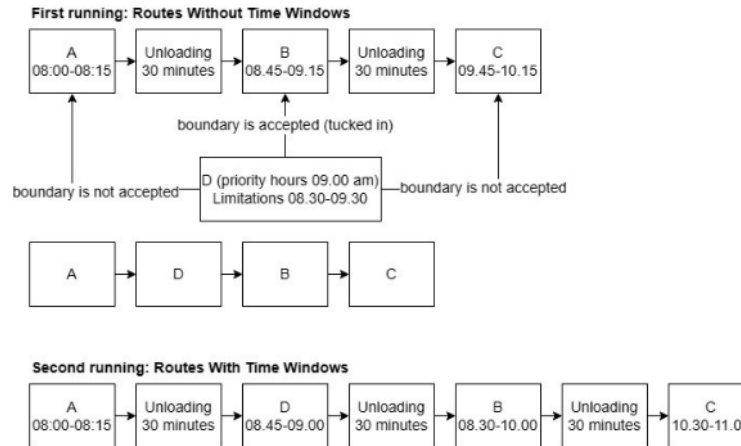


Fig 7. Algorithm of checking time windows

Table 4. The route with time windows

Priority	Area	Time Departure	Time Arrival	Duration (minutes)	Distance (km)
1	Pengampon	08:00:00	08:12:23	13	4.8
2	Gembong Sawah	08:42:23	08:47:45	6	1.6
3	Margomulyo	09:17:45	09:44:15	27	11.6
4	Kalimati	10:14:15	10:14:15	24	11.6

#### 4. Comparison of algorithm and manual calculations

After comparing the results of the daily routes for all trucks generated by the PSO algorithm on the system and manually by company employees for several days, there is a difference in the total duration and total distance traveled from the resulting route results. The results can be seen in Table 5.

The total time needed to produce the daily route for each of these trucks was recorded during the test. For route results that are generated manually, it will start to be calculated the first time the employee starts determining the route; the way the calculation is done is by using a stopwatch. Meanwhile, the time obtained from the algorithm on the system is obtained from the response time obtained when the system first calls the PSO algorithm function so that a table or agenda is formed for each truck.

It can be concluded that the total travel time duration of the trucks generated using the PSO algorithm is smaller than the routes generated manually by company employees. Likewise, the total mileage is smaller in the PSO algorithm than in the company's route results. The total time needed by the algorithm on the system to produce a shipment table for each truck is also much shorter than the manual route determination method by the company. In addition, by using the system, the number of trucks used is also more efficient.

Table 5. Durations and distances traveled on routes

Days	Manually		PSO Algorithm	
	Durations	Distances	Durations	Distances
Day One	40,910 seconds	182.48 m	37,702 seconds	166.79 m
Day Two	55,264 seconds	260.69 m	53,585 seconds	250.76 m

Day Three	48,389 seconds	180.99 m	41,040 seconds	158.02 m
Total	144,563 seconds	624.16 m	132,327 seconds	575.57 m

#### V. CONCLUSION

The system can provide estimates regarding the departure and arrival time at the destination with restrictions such as priority hours, payments, and time windows. It was found that the routes generated by the algorithm were shorter than manual calculation and used fewer trucks. The increase in response time is directly proportional to the increase in the number of destinations in a day

#### REFERENCES

- [1] P. Afrasyabi, M. S. Mesgari, E. M. El-kenawy, M. Kaveh, A. Ibrahim, N. Khodadadi, "A crossover-based multi-objective discrete particle swarm optimization model for solving multi-modal routing problems," *Decision Analytics Journal* vol. 9, 2023, <https://doi.org/10.1016/j.dajour.2023.100356>.
- [2] M.A. Hannan, M. Akhtar, R.A. Begum, H. Basri, A. Hussain, and E Scavino, "Capacitated vehicle-routing problem model for scheduled solid waste collection and route optimization using PSO algorithm," *Waste Management*, vol. 71, pp. 31-41, January 2018, ISSN 0956-053X, doi: <https://doi.org/10.1016/j.wasman.2017.10.019>.
- [3] A. Mussagulova, "A Particle Swarm Optimization for the Vehicle Routing Problem with Simultaneous Pickup and Delivery," Erasmus University Rotterdam, 2019. Retrieved from [https://thesis.eur.nl/pub/50115/Mussagulova\\_453776.pdf](https://thesis.eur.nl/pub/50115/Mussagulova_453776.pdf).
- [4] M. F. Ibrahim, M. M. Putri, D. Faritas, D. M. Utama, "An Improved Genetic Algorithm for Vehicle Routing Problem Pick-up and Delivery with Time Windows," *Jurnal Teknik Industri*, vol 22 no. 1, pp. 1-17, doi:<https://doi.org/10.22219/JTIUMM.Vol22.No1.1-1>
- [5] S C. Lo, "A particle swarm optimization approach to solve the vehicle routing problem with cross-docking and carbon emissions reduction in logistics management," *Logistics* vol.6 no. 3, 2022. Doi: <https://doi.org/10.3390/logistics6030062>.
- [6] Z. J. Peya, M. A. H. Akhand, T. Sultana, and M. M. H. Rahman, "Distance based Sweep Nearest Algorithm to Solve Capacitated Vehicle Routing Problem," *International Journal of Advanced Computer Science and Applications (IJACSA)*, vol. 10 no. 10, 2019, doi: <http://dx.doi.org/10.14569/IJACSA.2019.0101036>.
- [7] M. A. H. Akhand, Zahrul Jannat Peya and Kazuyuki Murase, "Capacitated Vehicle Routing Problem Solving using Adaptive Sweep and Velocity Tentative PSO," *International Journal of Advanced Computer Science and Applications (IJACSA)*, vol. 8 no. 12, 2017, doi: <http://dx.doi.org/10.14569/IJACSA.2017.081237>
- [8] Z. J. Peya, M. A. H. Akhand, K. Murase K, "Capacitated vehicle routing problem solving through adaptive sweep based clustering plus swarm intelligence based route optimization," *Orient.J. Comp. Sci. and Technol*, vol. 11 no. 2, 2018, doi: <http://dx.doi.org/10.13005/ojcs11.02.04>.
- [9] U. Dereci, M. E. Karabekmez, "The applications of multiple route optimization heuristics and meta-heuristic algorithms to solid waste transportation: A case study in Turkey," *Decision Analytics Journal* vol. 4, September 2022. Doi: <https://doi.org/10.1016/j.dajour.2022.100113>.
- [10] B. Su, Y. Lin, J. Wang, X. Quan, Z. Chang, and C. Rui, "Sewage treatment system for improving energy efficiency based on particle swarm optimization algorithm," *Energy Reports*, vol. 8, pp. 8701-8708, November 2022, doi:10.1016/j.egy.2022.06.053.
- [11] A. G. Gad, "Particle Swarm Optimization Algorithm and its applications: a systematic review," *Computational Methods in Engineering* 29, pp. 2531-2561, 2022, doi: <https://doi.org/10.1007/s11831-021-09694-4>.
- [12] M. A. Islam, Y. Gajpal, & T. Y. ElMekkawy, "Hybrid particle swarm optimization algorithm for solving the clustered vehicle routing problem," *Applied Soft Computing*, vol 110, October 2021. Doi:10.1016/j.asoc.2021.107655.
- [13] M. A. K. Azrag and T. A. A. Kadir, "Empirical Study of Segment Particle Swarm Optimization and Particle Swarm Optimization Algorithms," *International Journal of Advanced Computer Science and Applications (IJACSA)* vol. 10 no. 8, 2019, 10.14569/IJACSA.2019.0100862.
- [14] A. D. N. W. Susanto, H. Suparwito, "SVM-PSO Algorithm for Tweet Sentiment Analysis #BesokSenin," *Indonesia Journal of Information Systems (IJIS)* vol. 6 no. 1, 2023, <https://doi.org/10.24002/ijis.v6i1.7551>.
- [15] F. Kılıç, N. Uncu, "Modified swarm intelligence algorithms for the pharmacy duty scheduling problem," *Expert Systems with Applications* vol. 202, 2022. Doi: <https://doi.org/10.1016/j.eswa.2022.117246>.

# Swarm

---

## ORIGINALITY REPORT

---

20%

SIMILARITY INDEX

15%

INTERNET SOURCES

7%

PUBLICATIONS

4%

STUDENT PAPERS

---

## PRIMARY SOURCES

---

1	<a href="http://publication.petra.ac.id">publication.petra.ac.id</a> Internet Source	9%
2	Submitted to Arab Academy for Science, Technology & Maritime Transport CAIRO Student Paper	2%
3	<a href="http://umpir.ump.edu.my">umpir.ump.edu.my</a> Internet Source	1%
4	Submitted to Academic Library Consortium Student Paper	1%
5	Bingqin Su, Yuting Lin, Jian Wang, Xiaohui Quan, Zhankun Chang, Chuangxue Rui. "Sewage treatment system for improving energy efficiency based on particle swarm optimization algorithm", Energy Reports, 2022 Publication	1%
6	Javad Dogani, Farshad Khunjush. "Cloud Service Composition Using Genetic Algorithm and Particle Swarm Optimization", 2021 11th	1%

# International Conference on Computer Engineering and Knowledge (ICCKE), 2021

Publication

---

7	<a href="http://ejournal.umm.ac.id">ejournal.umm.ac.id</a> Internet Source	1 %
8	<a href="http://repository.petra.ac.id">repository.petra.ac.id</a> Internet Source	1 %
9	Submitted to The Robert Gordon University Student Paper	<1 %
10	Submitted to University of Western Australia Student Paper	<1 %
11	<a href="http://ebin.pub">ebin.pub</a> Internet Source	<1 %
12	<a href="http://pubsonline.informs.org">pubsonline.informs.org</a> Internet Source	<1 %
13	<a href="http://repository.unair.ac.id">repository.unair.ac.id</a> Internet Source	<1 %
14	Roy Orenge, M Christopher Maina, George N. Nyakoe. "Optimal Sizing and Placement of Solar Photovoltaic Based DGs in the IEEE 9 Bus System Using Particle Swarm Optimization Algorithm", 2018 IEEE PES/IAS PowerAfrica, 2018 Publication	<1 %
15	<a href="http://ijettjournal.org">ijettjournal.org</a> Internet Source	<1 %

---

---

16 "Proceedings of the 4th International Conference on Big Data Analytics for Cyber-Physical System in Smart City - Volume 1", Springer Science and Business Media LLC, 2023  
Publication <1 %

---

17 A. E. Rizzoli. "Ant colony optimization for real-world vehicle routing problems", Swarm Intelligence, 11/15/2007  
Publication <1 %

---

18 Communications in Computer and Information Science, 2015.  
Publication <1 %

---

19 Rolly Intan. "Mining Multidimensional Fuzzy Association Rules from a Normalized Database", 2008 International Conference on Convergence and Hybrid Information Technology, 08/2008  
Publication <1 %

---

20 [discovery.researcher.life](http://discovery.researcher.life)  
Internet Source <1 %

---

21 [josi.ft.unand.ac.id](http://josi.ft.unand.ac.id)  
Internet Source <1 %

---

22 [www.hindawi.com](http://www.hindawi.com)  
Internet Source <1 %

---

23

Parastoo Afrasyabi, Mohammad Saadi Mesgari, El-sayed M. El-kenawy, Mehrdad Kaveh, Abdelhameed Ibrahim, Nima Khodadadi. "A crossover-based multi-objective discrete particle swarm optimization model for solving multi-modal routing problems", Decision Analytics Journal, 2023

Publication

<1 %

---

Exclude quotes      On

Exclude matches      < 5 words

Exclude bibliography      On