Optimization Improved K-Means on Centroid Initialization process using Particle Swarm Optimization for Tsunami Prone Area Groupings

Gayatri Dwi Santika, Nadia Roosmalita Sari, M. Zoqi Sarwani, Wayan Firdaus Mahmudy

Abstract—Tsunami is a high wave caused by tectonic earthquakes, volcanic eruption or landslides in the ocean. Indonesia is one of the country that has thousands of islands. Lots of towns is a city on the banks or *water front city*. Indonesia becomes Tsunami prone areas. Tsunami can affect damage in various sectors, namely land degradation and infrastructure, environmental damage, fatalities, even the psychological impact on the victims themselves. Therefore, it takes a clustering of tsunami prone areas. The result of clustering can give information to the public to remain alert to the danger of tsunami. In addition, clustering of tsunami can be used by government to prepare policies in overcome the danger of tsunami. Improved K-Means is approach that proposed in this study to clustering the tsunami prone areas. In selecting the initial centroid must be done properly to produce a high accuracy. We proposed a method to determine the initial centroid appropriately, so that can increase the accuracy. The proposed method is Particle Swam Optimization (PSO). This study also uses comparison methods, such as K-Means, K-Means Improved, and K-Means Improved PSO. This study uses silhouette coefficient to test the accuracy of system. The result showed that the proposed method has higher accuracy than the comparison method. Silhouette coefficient generated at 0.99924223 with smaller computing time.

Index Terms—**Tsunami, classification, K-Means improved, particle swam optimization (PSO), silhouette coefficient**

I. INTRODUCTION

INDONESIA is one country that has thousands of small and large islands (about 13,677 islands) and has a very long coastline which is about 80 thousand kilometers. The history of the growth of cities and settlements in general occupy the site on the coast or along the rivers.

Lots of towns is a city on the banks or *water front city*[1]. Thus, Indonesia becomes prone areas for the tsunami. In December 2004 disaster Aceh which killed 130,000 people, 37,000 people missing, presumed dead, as well as 514,000 refugees and resulted in losses in the sector of environment, infrastructure, transport, communications, marine, energy and building amounting to US $ 4.3 billion[2]. Judging from these problems need his existing Tsunami anticipation in the area - Tsunami prone areas.

Tsunami is a high wave caused by tectonic earthquakes, volcanic eruptions or landslides in the ocean. Tsunami disaster causing damage to land, infrastructure, human fatalities, environmental damage, loss of property, and psychological impact. Tsunami disaster that ever happened in Pangandaran Beach has been one example of the Tsunami disaster that causes the loss of material andnon material on residents in coastal areas Kebumen[3]. By looking at the consequences of the Tsunami, it takes the Tsunami disaster response is one of them with the grouping of the areas potentially affected by the tsunami risk. With the potential risks in the area clustering, Tsunamis can give early containment is right for the area[4].

The classification of some data or objects into groups (clusters) is the concept of data mining. Data mining is an activity that includes the collection, use historical data to find regularities, patterns of relationships in large data sets. Output of data mining can be used for decision making in the future [5]. One of the techniques contained in data mining is clustering. Clustering is a method used to discover and classify data that is as similar as possible and different from the objects in

other clusters [6]. There are two methods of clustering, the hierarchical clustering and partitioning. Hierarchical clustering method itself consists of a complete linkage clustering, single linkage clustering, average linkage clustering and centroid linkage clustering. While the method of partitioning itself consists of a k-means and fuzzy k-means [5], [6].

Method of K-means clustering is a method that is the most simple and common. This is because the K-means has the ability to group data in large enough quantities to computation time is relatively quick and efficient [7],[8]. However, K-means have had weakness caused by the determination of the initial cluster centers. The results of the cluster is formed from the K-means method is very dependent on the initiation of the initial cluster center value given. This causes the cluster results in the form of local solutions that are optimal. To that end, the K-means collaborated with K-Means Improved methods for the determination of the initial cluster centers [9]. The difference between the K-Means algorithm with Improved K-means is the selection of initial centroid where K-Means is done randomly and the K-Means Improved calculation of the distance of the object with other objects that aims to get a better result.

Particle Swarm Optimization (PSO) is a type of evolutionary algorithm inspired by the flocks of birds [10]. Although the PSO method originally created for the completion of the optimization problem[11][12], the last few years PSO method has been applied to solve many problems associated with clustering. This is because the PSO is able to provide a more stable clustering results in the absence of dependence on initialization initial cluster centers (Hasan and Ramakrishnan, 2011).

Application of PSO on clustering performed by Van der Merwe and Engelbrech [13] who had the opportunity to produce a final value that is better than just applying PSO alone or K-Means alone. First clustering process carried out by PSO for determining the point of initial centroid followed by K-Means. Use of the initialization using PSO showed a high level of convergence and better than the use of K-Means.

An explanation of the problem and some research on tsunami, this study will implement methods Improved K-Means Clustering Encryption with PSO for determination in the centroid center grouping Tsunami prone areas in Indonesia.

II. RELATED WORK

Clustering is a strategy of grouping data in accordance with the characteristics of the data with other data. Previous research on the tsunami had already been done by Zezen, et al [4]. The result is some of the tsunami hazard level classification, namely high, medium, and low. However, C.45 algorithm can only be used to handle samples that can be stored as a whole and at the same time in memory.

Another study that raised the same issue is the research conducted by Oktariadi [14], using AHP (Analytical Hierarchy Process). Results from these studies is the level of tsunami hazard is divided into four levels, namely the danger zone, tsunami high, medium, and secure. AHP is known as a method capable of providing a hierarchical structure, taking into account the validity to the limit of tolerance and take into account the durability output sensitivity analysis of decision-making, but the AHP is a mathematical method without testing statistically and depend on the primary input that involves subjectivity of the experts, so it is difficult to determine weighting each criterion.

Research conducted by Pratt et al [15], namely the application of algorithms improve K-Means by using methods Invasive Weed Optimization (IWOKM). Performance of IWOKM method has been tested on some of the data and the results compared with the K-Means. From the tests, the result that the method IWOKM cluster capable of producing results better than K-Means.

To reduce the impact caused by the tsunami, then the required grouping vulnerable areas or potentially affected by the tsunami in order to provide appropriate measures for the region. Therefore do clustering in areas potentially affected by the tsunami. Methods commonly used to classify the data is K-Means. Several studies using the K-means. Several studies using the K-means as a method of research is Arai and Barathbah [16]and Ye et al[17].

The study [16]using the algorithm K-Means and methods appropriate for the initial grouping has high dimensional data. The results of this study indicate that the proposed method is effective and tends to be stable. A similar study on the development of data mining applications for telecommunications companies which are industries with intensive data Ye et al [17]Grouping customers can help analyze customer consumption accurately. K-Means implemented as customer segmentation method for Changzow Telecom in Jiangsu province. Research shows that the K-Means proven to be effective and successful in determining customer segmentation.

Customer segmentation was also scrutinized by Qin et al [18]on large data. However, in this case the researchers used as the Improved K-Means clustering customer data segmentation method. Improvement is based on the triangle inequality theorem. The results showed that the improved K-Means estimates the shorter time making it more suitable for large data. Implementation of improved K-Mean for centroid initial determination was made by Dongyang et al [19].K-Means algorithm is simple very sensitive to the initial determination of the center point. Selection of different initial center point results different clusters and cluster performance is strongly influenced by the initial center point. Improved K-Means can reduce the iterative process and improve the clustering performance.

A Study A. Ahmadyfard [20]discusses the performance of K-Means algorithm, PSO, and the

combination of K-Means with PSO. PSO algorithm shows better convergence during the initial phase of the global search, but in the process of searching around the global optimum will be very slow. Instead, the K-Means algorithm can achieve faster convergence to the optimal solution. At the same time, the accuracy converges for K-Means can be higher than PSO. The results of the study Ahmadyfard and H. Modares [20]showed that in the case of grouping data, the combination of K-Means algorithm-PSO is superior to a K-Means simple or PSO. Similar studies using K-Means combined with PSO to determine the optimal cluster centers. Research Naik et al [21]using a variety of multidimensional dataset.

In this study, we propose a clustering method that is a combination of the Particle Swam Optimization (PSO) algorithm Improved K-Means. In a previous study [21], improved K-Means algorithm combined with PSO algorithm is used as an approach in the case of a web document clustering. Web document is a large dataset. Therefore, the grouping is done effectively using partitional clustering algorithm on the approach proposed by Jaganathan and Jaiganesh [21]. K-Means clustering method is appropriate for handling large datasets more efficiently due to the time of execution. This study has used a variety of collections of text documents. The proposed method has better results compared with other test methods in this study

III. PROBLEM DESCRIPTION

The focus of this study is to determine the centroid early on Kmeans Improved method using Particle Swarm Optimization. In the initialization process starting point c, to determine the generalized centroid number of parameters automatically using PSO, after getting a C on the improved k-means, the value of c is inserted in the initialization process PSO as one individual. Furthermore, calculating the distance and silhouette coefficient resulting fitness as in Fig 1.
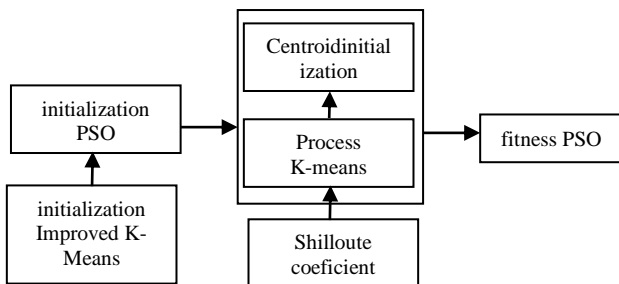

Fig 1. Iniziation centroid

a. K-Means

In K-Means conventional initial cluster center point selection is done randomly. Suppose the object has a variable p n making each grouped into k (k <n) cluster, where k is assumed a number of clusters. Variable object j i as xij (i = 1, ..., n; i = 1, ..., p). Distance Euclidean will be used as a measure of the difference in

this study despite the distance calculation may be used. Euclidean distance between objects i and j as follows [12]:

$$d_{ij} = \sqrt{\sum_{a=1}^{p}(x_{ia} - x_{ja})^2} \quad i= 1,......n; j= 1,.......,n \qquad (1)$$

1. Calculation the distance between each of all pairs of objects
2. Calculation vj for each j object , follow Equation 2 :

$$vj = \sum_{i=1}^{n} \frac{d_{ij}}{\sum_{l=1}^{n} d_{il}} \quad j= 1,...............,n \qquad (2)$$

3. Sorting vj in *ascending*. Take k object that has smallest value, a number of k. Use each object selected as initial cluster centers.
4. Determine the *cluster* initial in each object with the nearest cluster center obtained in step 3.
5. Calculation the sum of the distance of all objects to the center centroids.

Differences with the conventional K-Means K-Means Improved located in step 3. In the K-Means Improved data must in ascending sequence then determined the amount of his cluster. After that, the data is on each cluster summed and divided by a number of members of the cluster.

b. K-MEANS Centroid Optimization using Particle Swarm Optimization

PSO method was introduced by Kennedy and Eberhart [13]. PSO uses a set of cooperating particles, wherein each particle represents a candidate solution, to explore solutions that allow for optimization problems. Each particle is initialized at random or heuristic, then the particles - the particles are allowed to "fly". At each step of optimization, each particle is allowed to evaluate his ability and the ability of the particles in the vicinity. Each particle can store solutions that produce the best ability as one of the candidates the best solution for all the surrounding particles [13].

PSO is initialized with the manufacture of random matrices. The rows in a matrix called a particle. Such lines containing variable values. Each particle will move based on distance and speed. Particle renews speed (velocity) with equation 3 and its position is based on the local and global best solution to the equation 4 [13.]

$$Vit + 1 = \omega Vit + C1 * R1 * (Pit - Xit) + C2 * R2 * (Pgt - Xit)$$
$$(3)$$

$$Xit + 1 = Xit + Vit + 1$$
(4)

The variable i is the i-th particle in the herd, t is the number of iterations, Vi is the velocity of particles-I and Xi are variable vector particles (for example, the position vector) of the particle to-i in N-dimensional

problems. Pi is the best solution to local particle-i obtained, and Pg is the global best solution of all particles where Pi and Pg is obtained based on the best fitness values [13]. R1 and R2 is a random number between 0 and 1, ω is referred to as the particle load inner weight, C1 and C2 are two constants numbers, often referred to as cognitive confidence coefficient [13]. PSO clustering algorithm in case the data is preliminary determination of the centroid of each cluster formed in the K-Means.

IV. DATA

Tsunami Data obtained from *NOAA / WDS Global Historical Tsunami Database at NGDC: http://www.ngdc.noaa.gov/hasard/tsu_db.shtml.* The amount of data is entered at 116 records.

Table I. Tsunami data with 8 parameter

ID	CAUSE CODE	FOCAL DEPTH	PRIMARY MAGNITUDE	LATITUDE	LONGITUDE	MAXIMUM WATER HEIGHT	SOLOVIEV	TOTAL DEATHS
1	1	0	0	35	683	35.8	0	6
3	6	0	0	36.4	25.4	90	6	0
4	1	0	0	35	683	35.8	0	6
5	0	0	6	39.96	26.24	0	5	0
6	0	0	0	0	0	0	3	0
7	1	0	0	33.27	35.22	0	6	0
8	1	0	0	33.56	35.37	0	3	0
9	1	0	7	39.7	23.3	0	3	0
10	1	0	7.1	38.9	22.7	0	5	2500

The parameters used in this study using the 8 parameters that cause code (source tsunami), focal depth
(depth of the earthquake), primary magnitude (magnitude quake), latitude (location based on latitude), Longtitude (location based on longitude), maximum water height (wave height of water), Soloviev (the intensity of the tsunami), deaths (deaths). Data
Tsunami shown in Table I. Table I represents 10 data

sample of the data tsunami that is shown in this paper. Actually, this study used 100 data is an example of a data tsunami.

V. CENTROID INITIALIZATION

In the initialization phase, the centroid is determined randomly (random). The centroid is determined for each parameter for each cluster. This study This study divides the cluster into three clusters, low, medium, and high for each category of the tsunami with the vulnerability of low, medium, and high. Examples of centroid determination can be seen in Table II.

In Table III, known to each of the data of Table I calculated the distance to each cluster also includes cluster into. Data will only be entered in a cluster with

Table II. Centroid Initialization

Initialisasi	X1	X2	X3	X4	X5	X6	X7	X8
C1	0	35	683	6	0	0	683	0
C2	90	90	0	0	23.4	6	36.4	36.4
C3	35	35.8	0	6	0	683	35.8	1

the smallest distance value. Values d1, d2 and d3 is a distance value calculation results of each Tsunami with each cluster C1, C2, and C3. If among d1, d2 and d3 obtained distance value is the smallest is on d1, the tsunami will go on cluster c1. Examples of the Tsunami ID 1, after the calculation of distance with the reference

Table III. Cluster classification in iteration 1

ID	X1	X2	X3	X4	X5	X6	X7	X8	d1	d2	d3	Cluster
1	1	0	0	35	683	35.8	0	6	1184.42	1342.75	2011.605	C1
2	6	0	0	36.4	25.4	90	6	0	967.3408	943.7063	1493.959	C2
3	1	0	0	35	683	35.8	0	6	1184.42	1342.75	2011.605	C1
4	0	0	6	39.96	26.24	0	5	0	959.729	933.7298	1480.448	C2
5	0	0	0	0	0	0	3	0	964.4428	937.734	1486.233	C2
6	1	0	0	33.27	35.22	0	6	0	963.3422	937.8921	1485.643	C2
7	1	0	0	33.56	35.37	0	3	0	965.4665	940.0125	1488.301	C2
8	1	0	7	39.7	23.3	0	3	0	960.3544	934.1741	1481.044	C2
9	1	0	7.1	38.9	22.7	0	5	2500	2677.564	3614.225	4830.855	C1
10	3	0	7.3	38.25	22.25	0	5	0	958.6572	932.3783	1478.834	C2

Table IV. Cluster classification in iteration 1

ID	X1	X2	X3	X4	X5	X6	X7	X8	d1	d2	d3	Cluster
1	1	0	0	35	683	35.8	0	6	684.8595768	1314.655082	684.8595768	C1
2	6	0	0	36.4	25.4	90	6	0	100.708093	1164.098579	100.708093	C1
3	1	0	0	35	683	35.8	0	6	684.8595768	1314.655082	684.8595768	C1
4	0	0	6	39.96	26.24	0	5	0	48.43902559	1160.886934	48.43902559	C1
5	0	0	0	0	0	0	3	0	3	1161.89413	3	C1
6	1	0	0	33.27	35.22	0	6	0	48.82971739	1160.600245	48.82971739	C1
7	1	0	0	33.56	35.37	0	3	0	48.86011154	1160.626889	48.86011154	C1
8	1	0	7	39.7	23.3	0	3	0	46.66883328	1160.98061	46.66883328	C1
9	1	0	7.1	38.9	22.7	0	5	2500	2500.420947	1340.221608	2500.420947	C2
10	3	0	7.3	38.25	22.25	0	5	0	45.22626449	1160.972278	45.22626449	C1

C1 d1 values obtained at 1184.42, with a reference value obtained d2 C2 amounted to 1342.75, and with reference C3 d3 values obtained at 2011.605. Values d1 is the smallest value compared to the value of d2 and d3 then the Tsunami with ID 1 entry in the cluster C1. Placement of data on the corresponding cluster is temporary and can be changed after iteration. Examples of changes in the cluster because it is done iteration can be seen in Table IV. On the table it is known that the tsunami with ID 2, 4, 5, 6, 7, 8, and 10 were initially in the cluster C2 move into the cluster C1 while Tsunami with ID 9 which was originally located in the cluster C1 move into the cluster C2. Iteration will stop if there is no transfer of cluster members or forced to stop after a certain iteration predetermined. If we let the specified iteration will stop at the second iteration after iteration second grouping obtained from each cluster member. Based on the results in Table IV are obtained grouping of cluster members that can be seen in Table V.

Table V. Cluster result in iteration 2

CLUSTER	ID
C1	1, 2, 3 ,4, 5, 6, 7, 8, 10
C2	9
C3	

In this table it is known that the cluster C1 in getting members with ID 1, 2, 3, 4, 5, 6, 7, 8, and 10, and Reviewed obtained C2 member with ID 9, while the C3 does not get a member at all.

VI. CALCULATION THE SHILUETTE COEFICIENT

After the cluster is formed, then the validity of using the silhouette coefficient, see equation (5).

$$s(i) = \frac{b(i)-a(i)}{\max\{a(i),b(i)\}}$$
(5)

$a(i)$ is the average distance between objects in the cluster, and b (i) is the average minimum distance to objects in other clusters. Values silhouette coefficient ranges between -1 and 1. The larger the value, the better silhouette shows centroid obtained for each data. Because the C-Means Clustering Algorithm stochastic would require ten times the test so that it can obtain an average result that can be obtained. The results of these trials tenth can be seen in Table VI.

Table VI. The resultof siluette K-Means

Trial	Total Siluette Coeficient
1	0.9995651575472602
2	0.9988394145221908
3	0.9988394145221908
4	0.9997677497321314
5	0.9999834335834268
6	0.9984240803416191
7	0.99783375204334
8	0.9975132770951928
9	0.9999714520681304
10	0.9999714520681304
Shiluoette average	**0,999071**

Based on the table known to each total siluette of any trials conducted thus obtained an average total siluette that can be obtained from this algorithm is 0.999071.

To find out how effective and efficient PSO K-means, then do a simple comparison with the K-means clustering, improved K-means and PSO improved K-means. The result of the calculation silhouette coefficient on each cluster are shown in Table VII. Fluctuations in the silhouette in each cluster can be seen in Figure 2. Based on Figure 2 it can be seen that the first cluster has the best silhouette value.
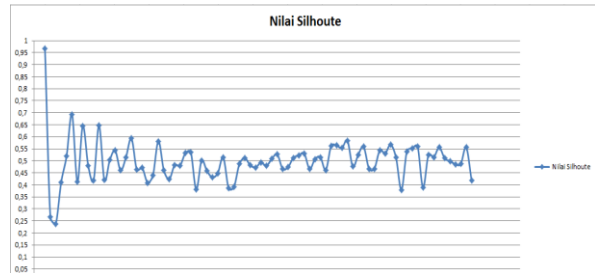

Figure 2. Result silhouette plot

VII. RESULT

This study uses multiple comparison methods to test the accuracy of the proposed methods. To test the accuracy of the system has been carried out using silhouette coefficient calculation technique.
In Table VII shows the results of the calculation

Table VII. The results of the comparison with the 4 Methods

Method	silhouette Coefficient result	Elapsed time
K-Means	0,999071	2 minutes 38 seconds
K-Means Improved	0.9988394	2 minutes 38 seconds
K-Means PSO	0,9999712	106 minutes 4 seconds
K-Means Improved PSO	0.99924223	47minutes24 seconds

Silhouette coefficient of each method. K-Means PSO obtained the highest accuracy with the silhouette coefficient value at 0.9999712. However, K-Means PSO hastime computation a very long time, which is 106 minutes. While the proposed method obtained the accuracy is not much different from the accuracy of the K-Means PSO. The proposed method obtained value at 0.99924223 with faster computation time. The proposed method gives the time execution faster than other comparison method. It is equal to 47 minutes with the highest accuracy. Based on the result, K-Means PSO is the best method.

VIII. CONCLUSION

Problems that related with clustering tsunami prone areas has been described in this study. Problems in determining the tsunami prone area is how to determine the initial centroid appropriately.
Combination of K-Means algorithm and the

improved PSO algorithm produces the best accuracy. The proposed method shows the optimal performance than the comparison method. There are K-Means, K-Means improved, and K-Means PSO. By uses silhouette coefficient, the proposed method produce value at 0.9992422 with the computation time is 47 minutes. Based on silhouette coefficient value and fast computation time, it can be concluded that the proposed method produces high accuracy.

This study is still need improvement, because the value of accuracy can still be improved. The higher silhouette coefficientthe more accurate the produced system. In the future study, we will make improvement by increasing silhouette coefficient value generated by the current study. The proposed method in future study is K-Means that combine with Genetic Algorithm (GA). GA is already widely used in problems related to optimization.

REFERENCES

[1]   D. Jokowinarno, "Mitigasi Bencana Tsunami Di Wilayah Pesisir Lampung(Tsunami Disaster Mitigation In The Coastal Lampung)" J. Rekayasa, vol. 15, no. 1, p. hal–13, 2011.

[2]   G. Strunz, J. Post, K. Zosseder, S. Wegscheider, M. Mück, T. Riedlinger, H. Mehl, S. Dech, J. Birkmann, N. Gebert, H. Harjono, H. Z. Anwar, Sumaryono, R. M. Khomarudin, and A. Muhari, "Tsunami risk assessment in Indonesia," *Nat. Hazards Earth Syst. Sci.*, vol. 11, no. 1, pp. 67–82, Jan. 2011.

[3]   F. Islam, S. Subiyanto, and L. M. Sabri, "Penentuan Resiko Dan Kerentanan Tsunami Di Kebumen Dengan Citra Alos,(Vulnerability Risk Assessment And Tsunami In Kebumen By Citra Alos)" J. Geod. Undip, vol. 3, no. 1, 2014.

[4]   Z. Abidin and A. Zezen, "Implementasi Algoritma C 4.5 untuk Menentukan Tingkat Bahaya Tsunami(Implementation Algorithm for Determining the Level C 4.5 Tsunami Hazard)," in Seminar Nasional Informatika 2011, 2011.

[5]   S. K. Yadav and S. Pal, "Data mining: A prediction for performance improvement of engineering students using classification," *ArXiv Prepr. ArXiv12033832*, 2012.

[6]   H.-S. Park and C.-H. Jun, "A simple and fast algorithm for K-medoids clustering," *Expert Syst. Appl.*, vol. 36, no. 2, pp. 3336–3341, Mar. 2009.

[7]   F. Yuan, Z.-H. Meng, H.-X. Zhang, and C.-R. Dong, "A new algorithm to get the initial centroids," in *Machine Learning and Cybernetics, 2004. Proceedings of 2004 International Conference on*, 2004, vol. 2, pp. 1191–1193.

[8]   Z. Huang, "Extensions to the k-means algorithm for clustering large data sets with categorical values," *Data Min. Knowl. Discov.*, vol. 2, no. 3, pp. 283–304, 1998.

[9]   T. Kanungo, D. M. Mount, N. S. Netanyahu, C. D. Piatko, R. Silverman, and A. Y. Wu, "An efficient k-means clustering algorithm: Analysis and implementation," *Pattern Anal. Mach. Intell. IEEE Trans. On*, vol. 24, no. 7, pp. 881–892, 2002.

[10] A. Ahmadyfard and H. Modares, "Combining PSO and k-means to enhance data clustering," in *Telecommunications, 2008. IST 2008. International Symposium on*, 2008, pp. 688–691.

[11] D. Novitasari, I. Cholissodin, and W. Mahmudy, "Hybridizing PSO With SA for Optimizing SVR Applied to Software Effort Estimation," *Telkomnika Telecommun. Comput. Electron. Control*, vol. 14, no. 1, pp. 245–253.

[12] W. Mahmudy, " 'Improved particle swarm optimization untuk menyelesaikan permasalahan part type selection dan machine loading pada flexible manufacturing system (FMS) (Improved particle swarm optimization for solving part type selection and machine loading problems in flexible manufacturing system)," *Konf. Nas. Sist. Inf. Univ. Airmadidi Minahasa Utara Sulawesi Utara*, pp. 1003–1008, Feb. 2015.

[13] D. W. van der Merwe and A. P. Engelbrecht, "Data clustering using particle swarm optimization," 2003, pp. 215–220.

[14] O. Oktariadi, "Penentuan Peringkat Bahaya Tsunami dengan Metode Analytical Hierarchy Process. Studi kasus: Wilayah Pesisir Kabupaten Sukabumi,(Determination of Tsunami Hazard Rating Method with Analytical Hierarchy Process. Case Study: Coastal Areas Sukabumi)" Indones. J. Geosci., vol. 4, no. 2, 2009.

[15] I. P. A. PRATAMA and A. Harjoko, "Penerapan Algoritma Invasive Weed Optimization Untuk Penentuan Titik Pusat Klaster Pada K-Means (Application Of Invasive Weed Optimization Algorithm For Determining The Center Point Of K-Means Cluster)," Universitas Gadjah Mada, 2014.

[16] K. Arai and A. R. Barakbah, "Hierarchical K-means: an algorithm for centroids initialization for K-means," *Rep. Fac. Sci. Eng.*, vol. 36, no. 1, pp. 25–31, 2007.

[17] L. Ye, C. Qiu-ru, X. Hai-xu, L. Yi-jun, and Y. Zhi-min, "Telecom customer segmentation with K-means clustering," in *Computer Science & Education (ICCSE), 2012 7th International Conference on*, 2012, pp. 648–651.

[18] X. Qin, S. Zheng, Y. Huang, and G. Deng, "Improved K-Means Algorithm and Application in Customer Segmentation," 2010, pp. 224–227.

[19] Dongyang Jiang, Wei Zheng, and Xiaoqing Lin, "Research on Selection of Initial Center Points Based on Improved K-means Algorithm," presented at the 2012 2nd International Conference on Computer Science and Network Technology, CHANGCHUN, CHINA, 2012.

[20] A. Ahmadyfard and H. Modares, "Combining PSO and k-means to enhance data clustering," in *Telecommunications, 2008. IST 2008. International Symposium on*, 2008, pp. 688–691.

[21] B. Naik, S. Swetanisha, D. K. Behera, S. Mahapatra, and B. K. Padhi, "Cooperative swarm based clustering algorithm based on PSO and k-means to find optimal cluster centroids," in *Computing and Communication Systems (NCCCS), 2012 National Conference on*, 2012, pp. 1–5.

[22] P. Jaganathan and S. Jaiganesh, "An improved K-means algorithm combined with particle swarm optimization approach for efficient web document clustering," in *Green Computing, Communication and Conservation of Energy (ICGCE), 2013 International Conference on*, 2013, pp. 772–776.