Journal of Information Technology and Computer Science
Volume 4, Number 1, June 2019, pp. 42-56
Journal Homepage: www.jitecs.ub.ac.id

Land Clustering for Potato Plants Using Hybrid Particle Swarm Optimization and K-Means Improved by Random Injection

Yudha Alif Auliya[1], Wayan Firdaus Mahmudy[2], Sudarto[3]

[1,2] Faculty of Computer Science, Computer Science, Brawijaya University, Malang, Indonesia
[3]Faculty of Agriculture, Brawijaya University, Malang, Indonesia
[1]yudha.alif7@gmail.com, [2]wayanfm@ub.ac.id,[3]sudarto_fpub@yahoo.co.id

Abstract. This research was conducted in Batu city, by classifying land based on land suitability for potato crops. Batu city is a hilly area with a high land slope so that there is a high potential for land degradation. Potato crop production is influenced by climate, suitability of planting land and treatment before harvest. Based on these problems, land mapping is needed so that it is easier for farmers to determine the optimal planting location for potato crops. The land mapping process is carried out using clustering techniques. The clustering process is carried out using 11 land suitability criteria for potato crops including average temperature, first month rainfall, second and third month rainfall, fourth month rainfall, drainage, soil texture, soil depth, Ph H2O, C-Organic, CEC and slope. The clustering results are 4 land suitability classes which are very suitable (S1), suitable (S2), quite suitable (S3) and not suitable (N). The clustering process is carried out using 5 different architectures namely K-Means, Particle swarm optimization (PSO), K-Means PSO, PSO K-Means, and Particle Swarm Optimization and K-Means (KCPSO) hybrids. The fitness value is calculated using the silhouette coefficient calculation. Architectural testing is done to get an architecture that has the highest fitness value. In this study a new approach was used to improve the accuracy of clustering results in the KCPSO architecture using the random injection method. Based on the test results, the KCPSO architecture obtained the biggest fitness values compared to the other fife clustering architectures. Testing the results of clustering is done by comparing the results of the KCPSO method with expert calculations.
***Keywords**:* Land Suitability, Clustering, K-Means and Particle Swarm Optimization

1 Introduction

Potatoes are an alternative carbohydrate sources that have lower glucose levels than rice so safe for diabetics. Potatoes is an excellent product for the agri plateau region because of high economic value and has a great potential in food diversification. Food diversification aims to strengthen national food resilience [1]. Government support in ensuring price and the smooth export-import potatoes also affect the process of the development of agribusiness potatoes.

Potatoes require certain requirements in order to be able to grow optimally consisting of the quality and characteristics of the land, the climate, topography, soils, hydrology and natural vegetation [2]. Potato cultivation oriented scale agribusiness should be planted in the appropriate area. To find suitable land and potential for the development of potato in Batu necessary land suitability evaluation. Land suitability evaluation is the process of estimating the level of land suitability by comparing the growth requirements, characteristics and quality of the land for a specific use with the potential resources available.

K-Means is a simple clustering algorithm that is without direction (unsupervised) [3]. The purpose of clustering is to group objects into k clusters or groups. In K-Means, the value of k must be determined in advance by considering the size of the dissimilarities in the group objects that exist. In addition to determining the number of clusters, other parameters that must be determined is the center of the cluster randomly. The better determination of the center of the cluster, the more accurate and faster process of grouping using the K-Means Clustering. However, since the center of the cluster is determined at random, then the level of accuracy is sometimes not good and frequent local optimum [4]. This can be prevented by using optimization algorithms to improve the first cluster center, so we get a better cluster center.

One of the optimization algorithms can be used to optimize the center of the cluster on the K-Means is Particle Swarm Optimization (PSO) [5]. Particle Swarm Optimization inspired by the natural phenomenon of bird behavior [6]. Hybrid K-Means PSO was made by Li, He, and Wen [7] to optimize the initial cluster centers so as to prevent local optimum. The same is also conducted by Cheng, Huang, and Chen [8] which proposes hybridization K-Means Clustering Particle Swarm Optimization (KCPSO). Hybrid K-Means PSO has also been used by Chiu, et.al [9] to overcome the problem of market segmentation.

The study will start by collecting the data fields that are required as input parameters are: average temperature, rainfall months to 1, rainfall second and third months, rainfall months to 4, humidity, drainage, soil texture, effective depth, Ph H2O, CEC and slope. There are three hundred datasets that will be grouped into four clusters, namely: very suitable (S1), appropriate (S2), is quite suitable (S3) and are not suitable (N). The execution was preceded by running Particle swarm optimization, then at a certain iteration carried out repairs on the center of the cluster. genes that describe the cluster center of each particle is included as initial cluster centers for K-Means algorithm. K-Means calculations performed using the initial cluster centers from PSO to obtain the new cluster center after Silhoutte Coefficient value converges. Center cluster has been updated by the K-Means to be put back on the next iteration of particles in Particle swarm optimization. Particle swarm optimization and K-Means is an integral process that does not run separately, but the output of particle swarm optimization process will be used as input for the K-Means, and vice versa. This iterative process continues until a predetermined number of iterations.

2 Land Suitability

Land suitability evaluation is the process of estimating the level of land suitability by comparing the growth requirements, characteristics and quality of the land for a specific use with the potential resources available [1]. Potato cultivation on suitable land, and the application of appropriate technology is expected to improve the productivity of potatoes in Kota Batu, with a low risk of crop failure. The main problem which leads to low productivity caused by low potato on land capacity and the condition of the hilly and mountainous relief with high slope. High slope land, causing erosion and land degradation.

In this study, the criteria of arable land will be augmented with external criteria that affect potato growth factors such as climate. Used 11 criteria land suitability average temperature, first month rainfall, second and third month rainfall, fourth month rainfall, drainage, soil texture, soil depth, Ph H2O, C-Organic, CEC and slope. Clustering of land based on land suitability criteria aims to classify the planting of land into four clusters, namely cluster that is very suitable (S1), appropriate (S2), is quite suitable (S3) and are not suitable (N). Clustering of land can improve the quality and quantity of potato plants produced by planting on land that has the highest land suitability criteria.

3 K-Means Clustering

As already described, K-Means is a simple clustering algorithm that is without direction (unsupervised). Suppose D is a dataset of n objects, and k is the number of clusters to be formed, the algorithm partitions organize these objects into the partition k (k n), where each partition describes a cluster. Each cluster is formed to optimize the partitioning criteria, such as the function of the difference based on distance, so objects in a cluster are similar, while objects in different clusters are not similar in terms of attributes dataset [10]. The equation for calculating the distance between the data on K-Means uses the formula Euclidiance Distance (D) shown in Equation (1).

$$D(X_2, X_1) = \sqrt{\sum_{j=1}^{p} (X_{2j} - X_{1j})^2} \tag{1}$$

Algorithm of K-Means is as follows [10]:
1. Inialization centroid K randomly.
2. Assign each object to the group with the nearest centroid. Use the Euclidean Distance formula to measure the minimum distance between data objects and each centroid using Equation (1)
3. Recalculate centroid vectors, using Equation (2)

$$mj = \frac{1}{nj} \sum_{\forall} data_p \in C_j data_p \tag{2}$$

mj denote the vector centroid of cluster j, n_j the number of vector data in cluster j, C_j c is part of the vector data of cluster j, and $data_p$ denote the vector data. Repeat steps 2 through centroid does not change anymore and in accordance with a predetermined maximum iteration.

4 Particle Swarm Optimization

Particle Swarm Optimization is a population-based algorithm inspired by the natural phenomena of the behavior of birds and fish. PSO is a meta-heuristic methods that optimize population-based problems with initializing a flock of birds randomly in the search space in which each bird is called a "particle" and the population of particles called "swarm" [11]. Particle swarm optimization is more flexible and have a balance between global and local searches on its search space. Particle swarm optimization is very efficient because it requires less computation than other evolutionary algorithms such as GA

The particles move in the iterative search space is calculated by displacement of the particle's position and velocity change formulas to find the best position globally. In the n-dimensional search space, position and velocity of a particle i at iteration t is

denoted by and vector $Vi(t) = (V_{i1}(t), V_{12}(t), \ldots, V_{in}(t))$. This solution was evaluated by a cost function for each particle at each stage of the algorithm to provide quantitative utility value of this solution. After that, record the best position of each particle based on the cost saved [11].
The best position of particle i previously visited during the current stage is represented by a vector $P_{i1}, P_{i2}, \ldots, P_{in}$ as a personal best. During this process, the position of all the particles that provide the best cost to the current stage is also recorded as the global best position denoted by . $G = (g_1, g_2, \ldots, g_n)$ Velocity structure and update the position illustrated in Figure 1. Each iteration consists of three movements: the first movement of a particle moving slightly toward the front to the previous direction at the same speed. In the second movement moves slightly toward the best position before. Finally, in the third movement, moving slightly towards global position. At each iteration, the speed and position of each particle is defined by Equation (3) and (4), respectively:

$$Vi(t) = \omega * V_i(t-1) + c_1\varphi_1\big(P_1 - X_1(t-1)\big) + c_2\varphi_2(G - X_1(t-1)) \quad (3)$$

$$X_i = X_i(t-1) + V_i(t) \quad (4)$$

α shows which introduces inertia preference for particles to continue moving in the same direction.

5 Silhouette Coefficient

Silhouette coefficient is a technique used to measure how well the location of objects in the cluster by providing a graphical representation of a short [12]. This method is a combination of cohesion and separation method [13]. In this study, Silhouette coefficient is used to calculate the fitness value of every particle in the process of PSO. Here are the stages of Silhouette coefficient calculation formula:

1. Calculate the average distance of a node i with all other nodes that are in a single cluster

$$a(i) = \frac{1}{[A]} \sum j \in_{A, j \neq 1} Cd(i,j) \quad (5)$$

2. Calculate the average distance from node i to all other nodes in the cluster and download the smallest value using Equation (6).

$$d(i,c) = \frac{1}{[A]} \sum j \in Cd(i,j) \quad (6)$$

d(i,C) = the average distance between nodes i with all the objects in other clusters or C, A $\neq$ C using Equation (7).

$$b(i) = minC \neq A(i,C) \quad (7)$$

3. Silhouette value of its coefficient is calculated using Equation (8) the following:

$$S(i) = \frac{b(i) - a(i)}{max\ (a(i), b(i))} \quad (8)$$

Average s (i) of all the data within a cluster indicates how close resemblance to the data within a cluster to determine how precise the data has been clusteredmanuscript.

6 Methodology

Clustering process the potato planted area will be optimized classification using multiple algorithms namely: K-Means algorithm, the K-Means PSO, PSO-Kmeans and hybrid PSO K-Means (KCPSO). In this paper do improvisation on KCPSO by applying random injection to obtain the optimal value of silhouette coefficient.

6.1 K-Means Clustering

The first step in the process of K-Means clustering is to select at random cluster center of the data that has been provided based on the number of attributes. Cluster center (centroid) with number 11 attributes shown in Table 1

Table 1: Determination of the initial cluster center

Centroid	Attribute	Attribute 2	…	Attribute	Attribute 1	Attribute 1
1	150	4	…	171	5	171
2	309	33	…	5	171	2
3	309	7	…	4	171	33
4	32	26	…	5	23	5

There are 4 centroid already established that the centroid 1 (C1), the centroid 2 (C2), centroid 3 (C3) and the centroid 4 (C4) then the next step is to calculate the proximity of each data in Table 1 with a fourth centroid. Results euclidean distance calculation shows that the shortest distance to the data 1 is the first centroid (C1). Determination of the cluster is determined based on the smallest of euclidience distance calculation. In the next iteration to amend centroid using equation (2). Centroid obtained after the K-Means process carried out continuously until the maximum iteration limit. In the K-Means method other than using a random method to determine the initial centroid there is a better method is to find the average value (average centroid). Average centroid method can minimize the occurrence of premature convergence.

6.2 Hybrid PSO K-Means (KCPSO)

KCPSO is the process of solving problems by using Particle swarm optimization in which there are stages of processing using the K-Means algorithm. PSO hybrid architecture K-Means has been used in previous studies [14]. The execution was preceded by using Particle swarm optimization, then on any given iteration period, for example at each iteration with a multiple of 5, carried out repairs on the center of the cluster. Repairs are meant genes which describes the cluster center of each particle is included as initial cluster centers for K-Means algorithm. Then the K-Means calculations performed using the initial cluster centers of the PSO, so that later obtained a new cluster center after Silhoutte Coefficient value converges. Center cluster has been updated by the K-Means to be put back on the next iteration of particles in Particle swarm optimization, ie iteration 6. This process is repeated until the maximum iteration.

6.3 Cycle of KCPSO Clustering

Step1 *I*nitialization particles containing centroid of each cluster and particle velocity

Step2 Determine the best particle and global best early

Step3 Start repetition and do:

Repetition number of particles and do:
Each vector data do:
Repetition number of particles and do:
Calculate the cost using the Silhouette coefficient.
(i) Calculated using the K-Means Clustering
(ii) Determine the cluster each vector with the K-Means Clustering
(iii) Calculate the cost using the Silhouette coefficient
(iv) Repetition of the last stops when the particles
(v) Update centroid using the update equation velocity and particle update
(vi) Update the best particle and global best

Step4 do random injection In particular iteration
Step5 Repetition stop when the maximum iteration

6.4 Initialization Particles

In the classification process, each particle in PSO represents a collection center cluster with each cluster consisting of many of the attributes. So if the cluster is assumed c={1, 2,….c} and attribute j={1,2,…j} each particle has a length *x=c x j*. nitialize the particles is shown in Figure 1

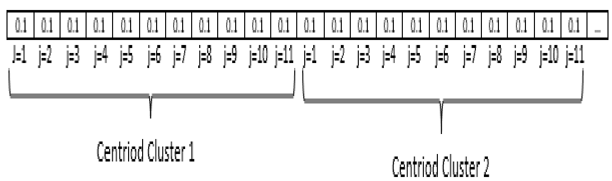

Fig. 1: Initialization a particles

In Figure. 2, eleven first column is the centroid for cluster 1, eleven next column shows the centroid of cluster 2 and so on. On each cluster consists of 11 attributes in the first column to attribute 1 and the second column for the attribute 2 until the eleventh column to attribute 11. This applies to the overall centroid..

The value for a random particles obtained with the minimum and maximum range of the overall data. Range used for each attribute are: 0.1 to 10 on the attribute 1 (j = 1), between 0.1 to 10 for attribute 2 (j = 2), between 0.1 up to 100 to attribute 3 (j = 3), between 1 up to 7 to attribute 4 (j = 4), between 1 to 200 for the attribute 5 (j = 5), between 1 and 7 to attribute 6 (j = 6), between 1 to 10 for the attribute 7 (j = 7). between 1 to 500 for attribute 8 (j = 8), between 1 to 500 for the attribute 9 (j = 9), between 1 to 200 for the attribute 10 (j = 10), between 1 and 50 for attribute 11 (j = 11).

6.5 Initialization of Velocity

The length of a long resembled a particle velocity. The length of the velocity v = c x j. The purpose of the update velocity is seeking a change to a more optimal centroid. The value of a velocity range between 0 and 1. The higher the value of velocity the more wide-term PSO algorithm in finding a solution.

Representation of velocity in Figure 3 shows that the value of 0.1 in the first column is the attribute value of velocity at 1 and 0.1 in the second column is the speed to attribute 2, as well as next to the attributes to 11. Velocity value in the first column up to eleven is the speed for the cluster centroid 1.

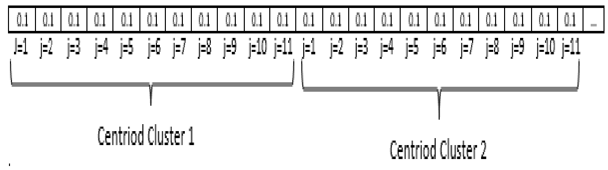

Fig. 2: Representation of velocity

This applies to the rated speed at 11 the next column to the next centroid. Inialisasi the initial velocity of each particle can be seen in Table 2

Table 2: velocity

P	*1*	*2*	...	...	...	*43*	*44*
1	0.1	0.1	…	…	…	0.1	0.1
2	0.1	0.1	…	…	…	0.1	0.1
3	0.1	0.1	…	…	…	0.1	0.1
...	...	…	…	…	…	…	…
20	0.1	0.1	…	…	…	0.1	0.1

6.6 Initialization of Population and Calculating Cost Value

Cost calculation starts with determining the clusters of each data in Table 1 by finding the smallest distance using the euclidean distance formula in equation (1) between the data and the centroid value of each cluster of particles that have been initialized. Determination of clusters of each data can be seen in Table 3

Table 3: Details of any data fields planting potatoes

No	*temperatures average*	*Rainfall month 1*	*...*	*KTK*	*Slope*	*Result Cluster*
1	23	389	…	42.7	2	2
2	23	389	…	33.4	5	3
3	23	389	…	33.4	5	3
4	23	389	…	32.4	10	4
…	…	…	…	…	…	…
197	26	540	…	42.7	35	1
198	25	389	…	39.9	10	4
199	25	389	…	42.7	10	4
200	25	502	…	42.8	35	4

A particle is considered to be the optimal solution by looking at the value of its cost. The smaller a value of cost, the better the particle used as a solution [15]. The value of cost at this clustering problem derived from the value of silhouette coefficient with the principle of greater value (closer to 1), the better a centroid is used as a clustering solution. Calculations derived from minimizing the cost function silhouette coefficient using Equation (9).

$$f(x) = \text{C} - f(x)' \tag{9}$$

f (x) is the value of cost derived from the final result silhouette coefficient is minimized. Value C as constants beginning to reduce the value of f (x) 'as the final value silhouette coefficient. In this study, the value of C to be used is 1 because the value f (x) 'only in the range of -1 to 1. Silhouette coefficient calculation process begins by calculating the average distance of a data (i) with other data within the cluster a (i). The next stage is to calculate the value of b (i) which is the average distance of the data (i) in the previous calculation with the data (i) in the other cluster. The number of clusters specified number two then the process of calculating the distance with only one other cluster. The last stage is the calculation of the value s (i) to obtain the results of their calculations in Table 4

Table 4: The Results of Each Stage Silhouette Coefficient

No	*a(i)*	*b(i)*	*s(i)*
1	33.98656	52.60452	0.35392
2	24.76588	51.52695	0.51936
3	24.33181	51.47338	0.52729
4	86.52450	75.59624	-0.12630
5	37.49010	36.59036	-0.02399
…	…	…	….
196	0.0	51.22376	1.0
197	0.0	197.05190	1.0
198	13.726087	126.52316	0.891513
199	6.38353	71.99079	0.911328
200	0.0	117.50893	1.0

Silhouette coefficient value is the average of the results of the calculation of s(i). Value silhouette coefficient should be minimized to obtain the value of cost, the cost value obtained is calculated using equation (9).
Table 5 is obtained from the population inialisasi inialisasi particles along with the final calculation of its cost. In a population consisting of a collection of particles with a population of some 20 has been set in advance calculations.

Table 5: Population

P	*1*	*2*	...	*44*	*Cost*
1	5.3	1.0	..	25.0	0.32613
2	5.3	1.3	..	26.0	0.31065
3	5.3	1.3	..	25.0	0.66170
…	...	...	..	...	...
20	5.3	1.0	..	23.0	0.28916

6.7 Determining Particle Best and Global best

Particle best determined at the beginning of the calculation. Particle best in a certain order the finest particles compared with particles in the previous iteration. Comparison of the particle each sequence is determined based on the cost that has been obtained. Particle best in early iterations will always be worth the same as the position of the particle at initialized which can be seen in Figure 4. In the early stages of each particle does not have perbanding particles.

5.3	3.5	28.0	…	…	272.0	23.0

Fig 3. Particle best

Global best obtained from comparing the overall particle of all the iterations that have been passed by value costnya. In early iterations, global best obtained from the particles in a population that has the smallest cost value. Based on the population in Figure 5, the particle with the smallest cost value and the best is to first particles so that the particle is defined as the global best.

5.66602	4.14446	29.85225	…	…	228.82156	23.98827

Fig. 4 Global best

Change the velocity and position of a particle centroid useful to find new, better and more optimal.

6.8 UpdateVelocity

Update velocity is useful to find the best position in the search space, iteratively speed undergo a change. By using equation (1), a change of pace would show up as a different value. Suppose it is known that w as inertia, c as a learning factor, and q is a random value with a range of 0 to 1. Particle global best and is best used as described in Section determines the best particle and the global best. So to look for a change of pace with the speed of the previous initial particle shown in Figure. 4, then obtained a change of pace on particles 1 and column 1 of 0.05.
$V_1(t) = (0.5 * 0.1) + 1 * 0.4 * (345 - 345) + 1 * 0.5 * (345 - 345)$
$V_1(t) = 0.05$

The calculation is performed on all of the particle velocity and on all the columns. So once calculated keselutruhan speed, it can be seen overall particle velocity changes as in Table 6

Table 6: Update velocity

P	*1*	*2*	*3*	...	*43*	*44*
1	0.59400	0.35343	0.82325	…	0.25090	0.25507
2	0.93051	0.94023	0.95435	…	0.66562	0.98412
3	0.93023	0.75790	0.41678	…	0.41678	0.41678
...	…	…	…	…	…	…
20	0.52731	-14.45	-19.45	…	0.41678	0.41678

6.9 Update Position of The Particle

After the update process velocity, continue the process of updating the particle positions to find the best particle position based on the position of the particle before and update velocity. The process of calculating the position of the particles was performed using equation (2). Suppose search for updated initial particle position shown in Figure 2, then the first column the resulting change in position is 345.05.
$X_1(t) = 345 + 0.05 = 345.0$. Similarly kecpatan changes, process changes the particle's position is also performed on all the columns and the entire particle.

6.10 Update Position of The Particle

The principle of determining the best particle and global best each iteration is to compare the values with those in the previous iteration. Determination of particle best by comparing the value of cost in the previous iteration best particles compared with the value of cost on changing the particle has been calculated. Values latest best particles are shown in Table 7.

Table 7. Update particel best

P	*1*	*2*	*...*	*44*	*Cost*
1	1.707	21.65	…	2.055	1.1482
2	7.297	1.641	…	6.812	1.1596
3	2.468	2.594	…	3.622	1.1476
…	…	…	…	…	...
20	3.719	3.9016	…	5.308	1.1629

Suppose the particle sequence 1 The value of cost on a first particle in Table 5 compared to the first cost value in Table 6 So the best particle in the first particle is a particle with the smallest cost value that is the first particle in Table 7. Determination of global best by comparing the value of cost on the overall position of the particles at each iteration. Best global search process is shown in Table 8

Table 8. Update global best based cost value

P	*Update Particle*	*particle Iteration – 1*
1	1.1482	1.1541
2	1.1596	1.1723
3	1.1476	1.1743
...	...	...
20	1.1629	1.1837

Based on global best search in Table 8 are obtained from the overall value of cost smallest particle that is equal to 1,154. Results of global best interpreted as the result of the best centroid. Centroid best results are shown in Figure 6

1.707	21.65	…	2.055

Fig 5. The Results of the Best Centroid

PSO clustering process continues until the maximum iteration limit. Having found the best results were obtained from the centroid global best, the best of the whole cluster of data were also obtained. One of the weaknesses in the PSO are often stuck in a local optimum solution, thus causing the value of cost generated experiencing premature convergence. To overcome these problems, a modification to the PSO that implement random injection process is usually done in the Genetic Algorithm to solve the same problem [15]. So that at certain iterations performed random particles results in order to prevent premature convergence.

7 Result

The first test was conducted to find optimal parameter combinations for the KCPSO architecture. The parameters tested were population size, number of iterations and percentage of hybrids. The maximum population test results can be seen in Figure 6

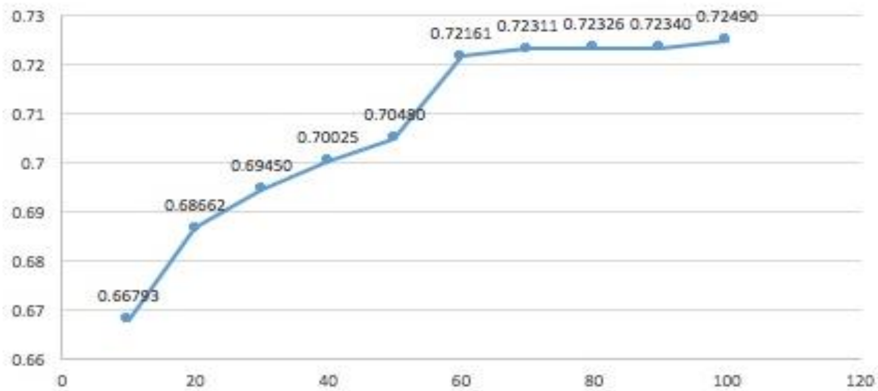

Fig. 6 Test results for the maximum population

Maximum population testing is done by setting the maximum population value of 10 by producing a fitness value of 0.66793. The resulting fitness value is not good because the optimum solution has not been achieved. The experiment was carried out by increasing the maximum number of populations to 70 and producing the best fitness value of 0.72311. Experiments continue to be made by adding a maximum population of up to 100 but the results are not significantly better. Based on these results the maximum population value is 70. The number of iteration test results can be seen in Figure 7

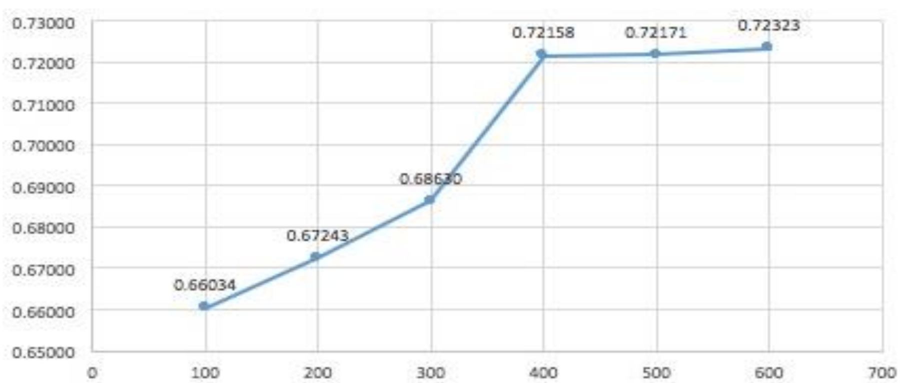

Fig. 7 Test results for number of iterations

Maximum iteration testing is done by setting the maximum iteration value of 100 by producing a fitness value of 0.66034. The resulting fitness value is not good because the optimum solution has not been achieved. The experiment was carried out by increasing the maximum iteration to 400 and producing the best fitness value of 0.72158. Experiments continue to be made by adding a maximum population of up to

600 but the results are not significantly better. Based on these results the maximum interation value is 400. The hybrid test results can be seen in Figure 8

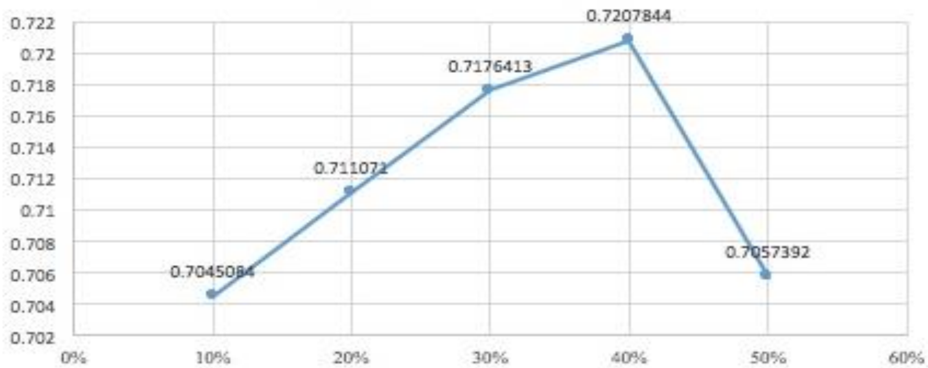

Fig. 8 Hybrid Test Results

Hybrid testing is done by setting the maximum iteration value of 10% by producing a fitness value of 0.7045084. The resulting fitness value is not good because the optimum solution has not been achieved. The experiment was carried out by increasing percentage of hybrid process to 40% and producing the best fitness value of 0.7207844. Experiments continue to be made by adding a maximum population of up to 50% but the fitness value gets smaller. based on these results the hybrid process will be carried out as much as 40% of the maximum number of iterations.

The second testing phase is by testing the random injection process in the KCPSO architecture. Application of random injection performed at specified intervals so that the computing time is not long but produces a better solution. Testing the parameters for the random injection process is carried out on the number of particles and the number of iterations. Testing the number of particles in the PSO is done to find the optimal combination of the number of particles each iteration that is processed using random injection. The results of testing the number of particles for the random injection process can be seen in Figure 9

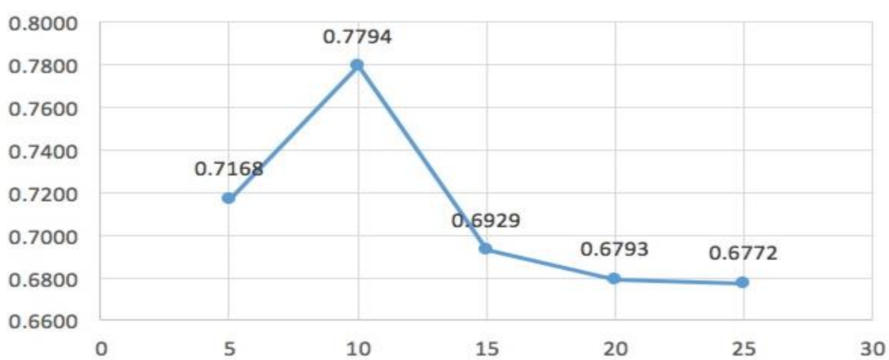

Fig 9. Results testing number of particles for random injection processes

Based on the test results it is known that in the random injection process experiments using 5 particles obtained a fitness value of 0.7168 and experienced a

significant increase when using 10 particles which amounted to 0.7794. But when testing using more than 10 particles the fitness value decreases because the resulting solution is faster convergence. Based on these results it is known that the optimal number of PSO particles for random injection is 10 particles.
The results of testing the number of iteration for the random injection process can be seen in Figure 10

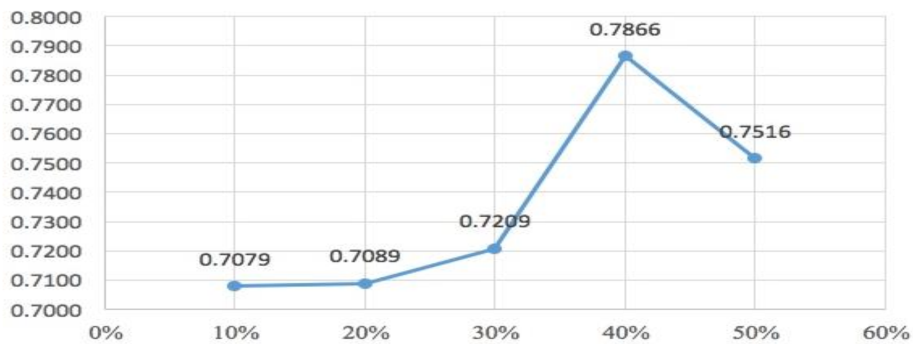

Fig. 10 Result Testing of Persentage Hybrid

Based on the test results it is known that the random injection process performed every 10% of the maximum iteration obtained a fitness value of 0.7079. The fitness value is getting better until the application of the random injection process every 40% of the maximum iteration with a value of 0.7866. When applying the number of iterations more than 40% of the fitness value decreases. Based on the results of the test it was concluded that the maximum number of iterations for the random injection process is 40% of the maximum iteration.

The process of clustering is done using a six algorithm to prove where a more accurate algorithm for clustering problems. 300 test data potatoes planted area run on four different methods. Each scenario method is run 20 times and then calculated the average fitness value. The results of the comparison of the four architectures are measured based on computational time which can be seen in Figure 11

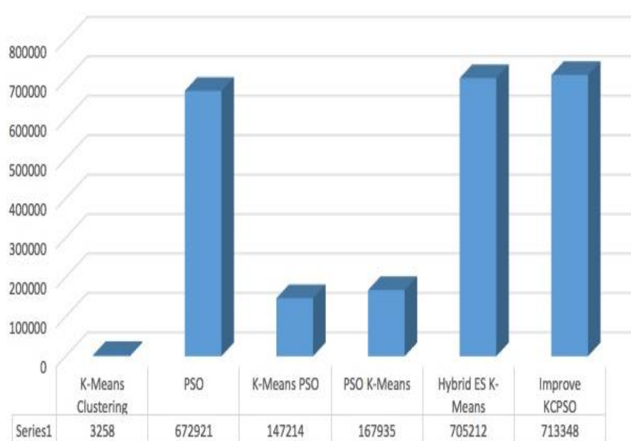

Fig 11. Results of Testing Computation Time

The results of the comparison of the six architectures are measured based on fitness values which can be seen in Figure 12

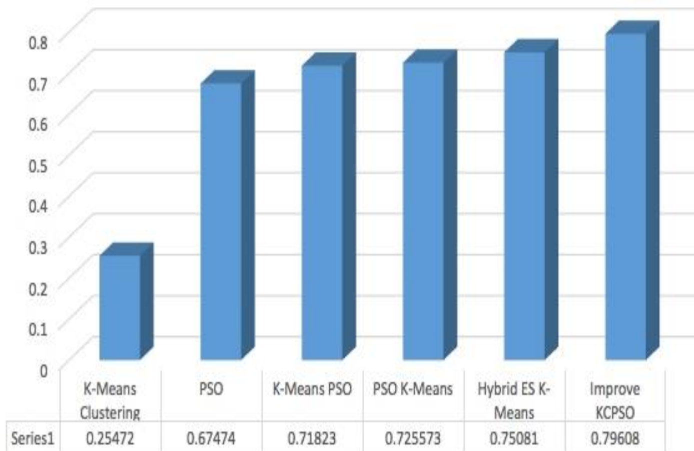

Fig 12. Comparison of Fitness Values

The author of this paper confirms that using the K-Means clustering course all the training data can be grouped into certain classes, but often the resulting solution is not a global optimum solution, but stuck in a local optimum value. K-Means method optimized using PSO can also solve the problem of clustering with silhouette coefficent values are better but still not optimal. The KCPSO architecture using random injection obtains the highest fitness value of 0.79608

After obtaining the best architecture, KCPSO, which is upgraded by random injection, is tested by clustering results by comparing the results of calculations performed by experts. Testing the results of clustering is done by comparing the results of the KCPSO improve with random injection with expert calculations using the 1976 FAO Framework guidelines. The results of 100 testing data compared to expert calculations obtained an accuracy of 86%.

7 Conclusion

Based on the testing that was done, it can be concluded that improve KCPSO can be used to solve problems of clustering with a high Coefficient Silhoutte value. KCPSO also proven to generate value Silhoutte Coefficient higher than the K-Means clustering algorithm, K-Means PSO and PSO K-Means. The use of random injection at a particular iteration interval can increase the value of coefficent silhouette. The computing time generated by the relatively high KCPSO improve so that in future studies need to be added to the testing interval iteration. This test is used to determine the number of iterations used in K-Means and PSO to obtain a balance between computation time and solutions provided. The resulting solution in the form of global optimum with little computing time. Testing the results of clustering is done by comparing the results of the KCPSO improve with random injection with expert calculations with an accuracy of 86%

References

1. M.F. Barcia, S.N. Muin and N. C. Deta, " Land Suitability correlation with temperature reference Suitability Red Potato Planting in Plain Medium Bengkulu ", *jur. Agroekotek*, vol.

2, no. 1, pp.21-26,2010

2. M. E. Djoemaijah, D. Dwiastuti dan D.Setyorini, " Assembled Test Potato Cultivation Technology Specific Location Highlands", *J. Pengkajian dan Pengembangan Teknologi Pertanian*. vol.2, pp. 104-110.
3. J. Karimov and M. Ozbayoglu, “Clustering Quality Improvement of k-means Using a Hybrid Evolutionary Model,” *Procedia Comput. Sci.*, vol. 61, pp. 38–45, 2015.
4. T. Niknam and B. Amiri, “An efficient hybrid approach based on PSO, ACO and k-means for cluster analysis,” *Appl. Soft Comput. J.*, vol. 10, no. 1, pp. 183–197, 2010.
5. R. J. Kuo, M. J. Wang, and T. W. Huang, “An application of particle swarm optimization algorithm to *clustering* analysis,” *Soft Comput.*, vol. 15, no. 3, pp. 533–542, 2011.
6. G. Armano and M. R. Farmani, “Multiobjective *clustering* analysis using particle swarm optimization,” *Expert Syst. Appl.*, vol. 55, pp. 184–193, 2016
7. H. Li, H. He, and Y. Wen, “Dynamic Particle Swarm Optimization and K-means *Clustering* Algorithm for Image Segmentation,” *Opt. - Int. J. Light Electron Opt.*, vol. 126, no. 24, pp. 4817–4822, 2015.
8. M. Y. Cheng, K. Y. Huang, and H. M. Chen, “K-means particle swarm optimization with embedded chaotic search for solving multidimensional problems,” *Appl. Math. Comput.*, vol. 219, no. 6, pp. 3091–3099, 2012
9. C.-Y. Chiu, Y.-F. Chen, I.-T. Kuo, and H. C. Ku, “An Intelligent Market Segmentation System Using K-Means and Particle Swarm Optimization,” *Expert Syst. Appl.*, vol. 36, no. 3, pp. 4558–4565, 2009.
10. J. Han and K. Micheline, *Data Mining: Concepts and Techniques*, Second Edi., no. Second Edition. Morgan Kaufmann Publishers, 2006.
11. A. Karami and M. Guerrero-Zapata, “A fuzzy anomaly detection system based on hybrid PSO-Kmeans algorithm in content-centric networks,” Neurocomputing, vol. 149, no. PC, pp. 1253–1269, 2015
12. L. Hidayat dan W. F. Mahmudy, “Grouping personality test result data bus driver using a genetic algorithm,” *Jurnal Teknologi Informasi dan Ilmu Komputer, vol. 3, no. 3, pp.163-168*
13. M. Anggara, H. Sujiani, and H. Nasution, “Selection of measure distance in k-means clustering for grouping members in alvaro fitness”, Jurnal Sistem dan Teknologi Informasi (JUSTIN)vol. 1, no. 1, pp. 1–6, 2016
14. Wahyuni, I., Auliya Y. A., Rahmi, A. dan Mahmudy, W. F. “Clustering based on the level of liquidity of bank customers using pso hybrid k-means”, *Jurnal Ilmiah Teknologi dan Informasi Asia* (JITIKA*)*, vol. 10, no. 1, pp. 20–30, 2016
15. W. F.Mahmudy, R. M. Marian and L. H. S. Loung, "Hybrid genetic algorithms for part type selection and machine loading problems with alternative production plans in flexible manufacturing system", *ECTI Transactions on Computer and Information Technology* (ECTI-CIT), vol. 8, no. 1, pp. 80-93