IJCCS (Indonesian Journal of Computing and Cybernetics Systems)
Vol.17, No.2, April 2023, pp. 195~206
ISSN (print): 1978-1520, ISSN (online): 2460-7258
DOI: 10.22146/ijccs.83195

Analyze the Clustering and Predicting Results of Palm Oil Production in Aceh Utara

Mutammimul Ula*[1], Gita Perdinanta[2], Rahmad Hidayat[3], Ilham Sahputa[4]
[1,2,4] Information System, Malikussaleh University, Lhokseumawe, Indonesia
[3]Department of Information Technology and Computer, Politeknik Negeri Lhokseumawe, Lhokseumawe, Indonesia
e-mail: ***[1]mutammimul@unimal.ac.id,** [2]gita.180180074@mhs.unimal.ac.id , [3]rahmad_hidayat@pnl.ac.id, [4]ilham.sahputra@unimal.ac.id

Abstrak

PT. Perkebunan Nusantara 1 merupakan perusahaan bidang produksi kelapa sawit dengan total luas lahan 1.144 Ha. Rumusan penelitian ini dapat menentukan cluster lahan produktif berdasarkan luas lahan, jumlah pohon, jumlah tahapan, dan produksi minyak sawit. Langkah-langkah metodologi antara lain data luas perkebunan dan data produksi kelapa sawit. Penelitian ini dapat membandingkan Cluster C-means dan K-means dalam melihat lahan produktif. Sedangkan untuk prediksi hasil produksi menggunakan algoritma Backpropagation Neural Network (BPNN) dan Fuzzy time series. Hasil cluster data produksi minyak sawit Cot girek periode 2019-2022 dari Januari hingga Desember adalah 1.365.530, sedangkan pada tahun 2022 mencapai 1.768.720. Analisis cluster lahan seluas 1.144 hektar yang diperoleh 800,4 hektar lahan produktif dan 343,6 hektar lahan kurang efektif. Hasil model clustering C-means lebih dari K-meas dengan iterasi yang lebih singkat sedangkan untuk prediksi memiliki tingkat akurasi sebesar 90,77%. Hasil perbadingan peramalan dapat melihat tingkat akurasi fuzzy time series adalah sebesar 81,27%. Selanjutnya hasil penelitian ini dapat dijadikan rekomendasi bagi perusahaan dalam analisis analisis pengelompokan lahan produktif dan hasil prakiraan dari lahan-lahan tersebut.

Kata Kunci*— Pengelompokan, K-means, C-Means, Prediksi, Backpropagation*

Abstract

PT. Perkebunan Nusantara 1 is engaged in oil palm production with a total land area of 1,144 Ha. The formulation of this research can determine productive land clusters based on land area, number of trees, number of stages, and palm oil production. Methodological steps include plantation area data and oil palm production data. This study can compare the C-means and K-means groups. As for predictions using the Backpropagation Neural Network (BPNN) algorithm and Fuzzy time series for production results. The results of grouping Cot girek palm oil production data for the 2019-2022 period from January to December were 1,365,530, while in 2022 it reached 1,768,720. The analysis used a land grouping method of 1,144 hectares, which resulted in 800.4 hectares of productive land and 343.6 hectares of less effective land. The results of the C-menas clustering model are more than K-meas with shorter iterations while for predictions it has an accuracy rate of 90.77%. As a comparison, the level of accuracy of the fuzzy time series is 81.27%. The results of this study can be used as recommendations for companies in the analysis of productive land grouping analysis and forecast results from these lands.

Keywords— *Clustering, K-means, C-Means, Prediction, Backpropagation*

Received March 16th,2023; Revised April 28th, 2023; Accepted April 30th, 2023

1. INTRODUCTION

Indonesia is the largest producer of palm oil in the world. The Indonesian Ministry of Agriculture (Kementan) estimates that there are around 15.08 million hectares (Ha) of oil palm in 2021 [1]. Indonesia is one of the producers of Crude Palm Oil (CPO) of 49.7 million tons in 2021 [2], [3]. PT Cot Girek PKS is one of the branches of PT Perkebunan Nusantara 1 (Persero) which was founded in 1991 [4]. The processing capacity of the Cot Girek palm oil mill is estimated at 30 tonnes/hour of fresh fruit bunches (FFB) and can be increased to a maximum of 45 tonnes/hour.

The problem that must be considered by the company is how to find out the productive land clusters and less productive land. Furthermore, there is no prediction of palm oil production in the next few years, this prediction is important for companies to be able to maximize the performance. Whereas, clustering and predictions can suggest a quick policy directions to stabilize the palm oil production and the ability to improve in every year.

Cluster analysis can be used to group the productive and non-productive land production data by using a model that includes statistics, artificial intelligence, machine learning and predictions. Analysis of the clustering model can classify the productive palm oil lands, while the C-Means Clustering and K-means are included in the model. The results of the two models can be used as the best accuracy value in clustering [5],[6].

The accuracy of the research using C-Means and K-means Clustering in grouping of productive land while the prediction using the Neural Network Backpropagation and Average Fuzzy Time Series models serves as a prediction in determining of future palm oil production [7]. To determine the accuracy of the two models, clustering can be considered by the two models, namely K-Means and C-Means while predictions were Backpropagation Neural Network (BPNN) and Fuzzy Time Series [8],[9].

In this research, productive land clustering makes it easier for companies to determine the productive land and less productive land [10]. Furthermore, the company needs a prediction to find out how far the productive and less productive land are in producing palm oil. This can reduce the loss for the company in the future. The company can decide to make observations on land that is less productive to be considered in further development in the future [11]. The importance of the prediction in palm oil production is to find out whether palm oil production reaches the target or not. If it does not match the prediction, the company can evaluate it according to the expected target [12]. In this research using the clustering and prediction model was carried out to cluster the land and predicting the palm oil production according to the predicted target [13], [14], [15].

Related research according to Noviar, et al in 2019, the result of their research is to find out how many clusters are optimal and which group for each productive land cluster is the most dominant of these clusters. The variables used in this research are type of oil palm, year of planting, area, and average production. There are three categories of the clustering results, namely high productivity (C1), medium (C2), and low (C3). The data is obtained from the monthly data summary in the company [16]. Subsequent research on land clustering based on palm oil varieties with the Fuzzy C-Means model. The results obtained 3 clusters formed and the grouping model of 9 repetitions for land suitability based on the variety of the palm oil plant [17].

The next research by using the Backpropagation neural network model in predicting the future palm oil production. The result of this research achieve an accuracy of 82% of the Backpropagation model. This model can be used in predicting the productivity of palm oil in companies so that the results of the predictions will become an information and suggestion in making the right policies so that the productivity of palm oil in these plantations is relatively stable and able to increase every year [13].

Subsequent research on land clustering can predict Crude Palm Oil (CPO) commodity prices with a time series model. The test results were carried out using MAPE to predict CPO prices with an accuracy of 0.01781302%. Time series testing on CPO price data with the best accuracy is calculated using MAPE without involving other model combinations [18]. The prediction model can also predict the production of palm oil in companies which varies each period. The results of the research on predictions the palm oil production in this company vary each period, so the problem that often occurs is insufficient supply and demand. The results with the FTS model with a MAPE value of 9.05895% [19].

Based on these problems, this study examines the results of comparing productive and unproductive oil palm areas using the C-Means and K-means clustering method in Cot girek PKS plantations. The two methods can see from the results of each grouping which is better in seeing productive and non-productive land. Next, compare the results of forecasting methods for next year's palm oil production, whether it reaches the target set or not, with the BPNN and FTS methods. This research can see the results of a comparison of effective methods in grouping and forecasting.

The benefits of research in the academic field can find out effective data mining cluster patterns and artificial intelligence, where the cluster method can classify productive oil palm lands and not based on production levels. Then the benefits in the academic field of forecasting can be helped by the campus/academic to find out how to classify oil palm lands based on productivity levels and can help companies see plantation productivity in the future.

2. METHODS

2.1 Dataset Processing

The dataset used is the Cot Girek PKS plantation from 2019 to 2021 on palm oil production. The collected datasets were processed using C-Means and K-Means methods for clustering productive land as shown in Table 1. Meanwhile the BPNN and FTS models for predicting palm oil production are shown in Table 2. The following variables used in this research include block area, land area (ha), number of trees, palm bunches and number of oil production stages. Table 1 is a dataset in clustering the productive and less productive land as follows:

Table 1 Production Dataset

No	Afd	Block	Areas (Ha)	NoT	Palm Bunches	Production
1	IV	1212	19	2603	306	5060
2	IV	1214	10,5	1438	164	2830
3	IV	1212	3	408	74	1220
4	IV	1213	18	1805	278	4640
5	IV	1214	7	454	111	1830
6	IV	1313	10	1239	164	2690
85	IX	2027	10	1317	401	6130
87	X	1827	2	270	78	1240
…	…	…	…	…	…	…
96	X	31 W	2	285	85	1330
97	X	28 X	16,5	2381	502	7810
99	X	1229	5,4	410	150	2770
100	X	1230	9	670	170	3720

The data in Table 2 used for predicting using BPNN and Fuzzy time series as follows:

Table 2 Predicting Dataset

Month/Year	2019	2020	2021	Target
Januari	135.430	128.530	131.350	131.770
Februari	110.440	121.890	154.510	128.947
Maret	69.140	77.770	156.250	101.053
April	131.610	111.440	158.230	133.760
Mei	145.190	63.810	127.830	112.277
Juni	137.000	109.890	180.720	142.537
Juli	97.950	123.060	96.890	105.967
Agustus	113.660	120.750	155.320	129.910
September	143.200	126.760	165.290	145.083
Oktober	70.980	105.990	143.700	106.890
November	94.970	126.030	158.000	126.333
Desember	115.960	81.010	140.630	112.533

2.2 C-Means Clustering

Fuzzy C-Means is a grouping of data based on the degree of membership in each data [21]. FCM data can be a member of a class or cluster that has membership degrees varying between 0 and 1. The formula for determining the centroid value is as follows:

$$V_{kj}=\frac{\sum_{i-1}^{n_k} X_{ij}}{n_k} \quad (1)$$

Where V_kj is centroid, k is a cluster index, j is variable index and X_ij object value of -1 in the cluster for variable -j.

2.3 K-Means Clustering

The clustering step in K-Means is to determine the k value as the number of clusters [22]. To calculate the distance of each input data to each centroid is using the Euclidean distance until the closest distance is found from each data to the centroid [23].

$$D_{(i,j)} \sqrt{(X_{1i} - X_{1j})^2 + X_{2i} - X_{2j})^2 + \cdots + (X_{ki} - X_{kj})^2} \quad (3)$$

Where D_((i,j)) is the distance of data -i to centroid j, X_ki is data of -i in variable k, and X_kj is centroid j of variable k. The next step is to find the centroid value of each of the k clusters. The cluster center point (centroid) is the average of all points in the cluster [24].

2.4 Backpropagation Neural Network

BPNN's steps include feedforward, backward and weight updates. The weights are randomly reset at intervals of -1 to 1 or 0 to 1. The activation functions that can be used are binary sigmoid functions between 0 and 1 or bipolar sigmoid functions between -1 and 1 [25].

There are several methods to measure the predictive models including mean absolute error (MAE), mean square error (MSE), root mean square error (RMSE), and normalized mean square error (NMSE). This research used MSE in measuring prediction accuracy. The following is the MSE formula [26].

$$MSE = \frac{1}{M} \sum_{t=1}^{M} (x_t - x_t)^2 \quad (4)$$

Where, x_t is the data value x_t - x_t is the prediction and M is error rate. In this research, the time series data is normalized to the order [0,1] with the following formula:

$$\bar{x} = \frac{(x - x_{min})}{(x_{max} - x_{min})} \quad (5)$$

Where x is a normalize time series data, -x is real time series; x_max is maximum time series and x_min is the minimum time series. After that, the denormalization process is carried out to return to the original data [27].

The forward propagation phase is any signal that enters the neuron from the hidden layer (z). The following is the input signal and at each weight using the following formula.

$$zin_j = vo_j + \sum_{k=0}^{n} x_i \, v_{ij} \quad (6)$$

The backward propagation phase is to calculate the output error correction factor, the output error correction is the resulting error for the resulting unit error y_k in the following formula.

$$\delta_k = (t_k - y_k)\, y_k\, (1 - y_k) \quad (7)$$

After following the steps to find the forward propagation phase, the next step is to calculate the change in weight w_jk with the following formula.

$$w_{jk}(\text{new}) = w_{jk}(\text{old}) + \Delta w_{jk} \quad (8)$$

Where w_jk old is the weighting value of w_jk.

2.5 Fuzzy Time Series

Fuzzy time series is a data prediction method that uses fuzzy principles as its basis. Where the fuzzy set is defined as a class of numbers with fuzzy boundaries. If the universe of discourse (U) is the set of universes, U = [u_1,u_2,…,u_p], then a fuzzy set of A_(1) from U with the degree of membership is generally expressed by the following formula:

$$A_i = \mu_{Ai}(u_1)/u_1 + \ldots + \mu_{Ap}(u_p) / u_p \quad (9)$$

Where A_i= 〖(μ_Ai) is the degree of membership of u_1 to A_i, where μ_Ai (u_(i)

)∈[0,1] and 9≤i ≤p. The degree of membership value of μ_Ai (u_1) define as follow.

$$\mu_{Ai}(u_i) = \begin{cases} 1 & if\ i = j \\ 0{,}5 & if\ i = j-1\ or\ j+1 \\ 0 & else \end{cases} \tag{10}$$

This can be described by rules, if X_t is in u_i, then the degree of membership for u_i is 1, and u_(i+1) is 0,5 and if not u_i and u_(i+1), then the value is 0 [28].

The next step is to determine the number of clusters of productive land. As well as to determine the production predictions for the following year with the backpropagation method. The following is the flow of the research.

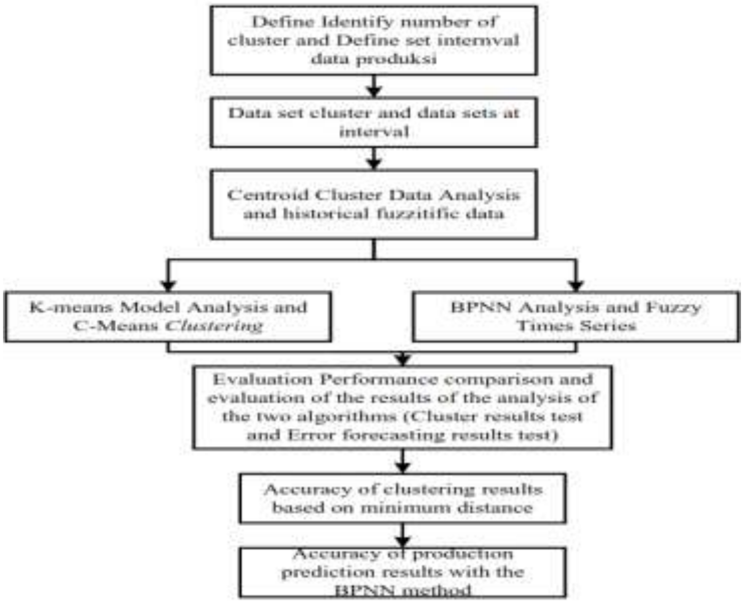


Figure 1. Research Flow

The initial stage of the research methodology is to collect the datasets at PT Perkebunan Nusantara I. These datasets will be processed to determine clustering and prediction results. Next, determine the effect of the results of each cluster, where the cluster results are used for the analysis of clustering productive and less productive land. The cluster are in the form of the best, moderate, and bad clusters. Next, determine the value of the variables in the previous year's production and predict the following year's production. The final step is to compare the accuracy of each method to determine the best way for clustering and prediction where the methods compared are the c-means and k-means clustering for the clustering process and backpropagation neural networks and fuzzy time series for the prediction models to determine the level of the prediction accuracy [20].

3. RESULTS AND DISCUSSION

3.1 The result of C-Means and K-Means

Clustering and predictions in finding the productive performance of palm oil in cot girek mills are shown in the following steps:

3.1.1 Manual Model of C-Means

A. Determine the random value

In calculating C-Means, the first step is to determine the random values where to generate random values the value cannot be more than 1 (<1). The following is the results of the random value in the cluster as shown in Table 3.

Table 3. Random Value Results

$\mu1$	$\mu2$	$\mu3$	$\mu4$
0,09	0,04	0,16	0,01
0,04	0,09	0,04	0,09
0,25	0,01	0,04	0,04
0,01	0,09	0,09	0,09
0,16	0,04	0,09	0,01
0,04	0,09	0,04	0,09
0,54	0,41	0,34	0,41

The next step is to determine the square measure for each cluster divided by the square measure of the total, as shown in the following table.

Table 4. Centroid

uik	xij	uik^2	$uik * xij$
23944,72	206,55	24295,56	117017,04
25700,66	152,44	17429,63	77146,10
25944,76	186,75	20858,76	100964,41
24364,34	146,73	17338,76	73341,95

The steps in determining the cluster center point are as follows:

P1,1 = $\frac{12930,15}{0,54}$ = 23944,72

P1,2 = $\frac{111,54}{0,54}$ = 206,55

P1,3 = $\frac{10537,27}{0,41}$ = 24295,56

P1,4 = $\frac{63189,20}{0,54}$ = 117017,04

B. Objective Function Value

The next step is to determine the value of the objective function by means of the cluster center multiplied by the squared value as shown in the following table.

Table 5. Objective Function Value

Xi^w	Xj^w	Xk^w	$Xi^w + Xj^w Xi^w$
6973,3	2311,2	6972,6	2311,75
1994,1	0,00	1994,3	0,00
0,66	1994,8	0,50	1994,8
9787,8	1951,1	9787,4	1951,9
4240,5	4581,8	4240,5	4581,2

The next step is to subtract the epsilon value from the objective function value with conditions [P_t-P_(t-1)] < epilson because the objective function in the 1st iteration is 160293,53 > 0,05. The next step is to obtain the objective function value which is smaller than the epsilon value.

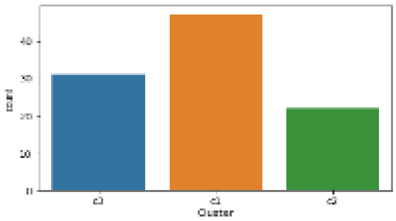
The following is the results of clusters based on the degree of membership in the last iteration to determine the best clusters, good clusters, and poor clusters.

C. C-Means Clustering Result

Table 6. C-Means Clustering Result

Block	Areas	NoT	Palm Bunches	Production	Cluster
1214	10,5	1438	164	2830	C1
1212	3	408	74	1220	C1
1214	7	454	111	1830	C1
1313	10	1239	164	2690	C1
1420	9	1164	333	5870	C2
1825	11	1573	261	4300	C2
1314	25	3429	358	5990	C3
.....	....	......	.....	......	......
2128	19	1196	870	16190	C3

Then we get clusters based on the dataset with the epsilon value that has been obtained into three clusters, namely best cluster (C1), good cluster (C2) and poor cluster (C3).



The following is the distribution of the C-Means clustering processing, it can be seen that the blue color is the best cluster value with the x label being the number of principals and the y label being the land area with the density of each data approaching the cluster points.

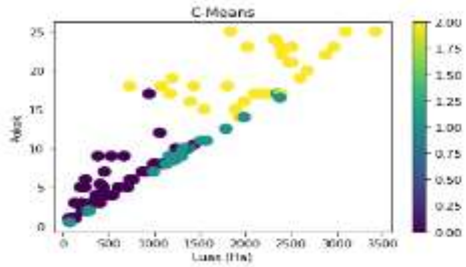


Figure 2. The distribution of C-Means Clustering

3.1.2 Manual model of K-Means

The first stage for clustering using the K-Means method is to determine the centroid center of each cluster. After determining the cluster center, the next step is the same as the C-Means method, namely finding the cluster center value and the squared value. Where the K-Means method cluster table is as follows.

Table 7. K-Means Clustering Results

No.	C1	C2	C3	Distance	*Cluster*
1	2.349	5.060	10.120	2.349	C1
2	1.378	2.830	5.660	1.378	C1
3	1.095	1.220	2.440	1.095	C1
4	1.396	820	1.640	820	C2
5	2.672	2.669	2.648	2.648	C3
6	1.465	1.462	1.440	1.440	C3
7	379	402	382	379	C1
....	.....	.....	.....	......	.....
100	2.393	2.200	4.400	2.200	C1

The following is the distribution of clustering results on the k-means method.

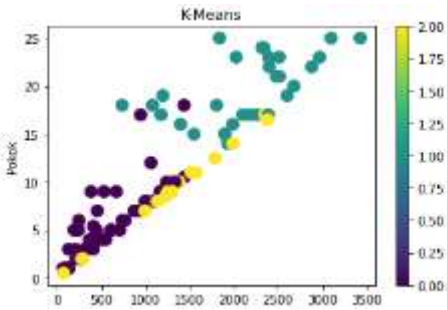


Figure 3. The distribution of K-Means Clustering

Where the distribution in the K-Means method displays the distance from each cluster, it can be seen that the blue color is the best cluster, the yellow is medium cluster and the purple is poor cluster.

3.1.3 The Comparison of Accuracy of C-Means and K-Means

Based on the clustering results obtained for clustering the productive land on Cot Girek PKS plantation. The analysis of the number of clusters using the Fuzzy C-Means method are better than the K-Means method in terms of clusters. It can be seen from the epsilon value where the c-means method is closer to the epsilon value than the K-Means method. The best cluster results are in cluster C1 with a value of 47%, the medium cluster is 22% and the poor cluster C3 is 31%.

3.2 Prediction Results of Backpropagation Neural Network and Fuzzy Time Series

3.2.1 Manual model of Backpropagation Neural Network

The first step is to find the minimum and maximum values of the dataset. The following are the calculation steps using the backpropagation neural network method.

A. Min Max Value

Data normalization used 12 data per year, from 2019 to 2021. Data is normalized to find the maximum and minimum value parameters.

Table 8. Min Max Value

	X1	X2	X3	T
Min	69.140	96.810	96.890	100.000
Max	145.190	128.530	180.720	200.00

The next step is to calculate the data normalization value by looking for the forward propagation, backward propagation and flashback propagation values as follows:

$X1 = \frac{135,430 - 69,140}{145,190 - 69,140}(0,8) + 0,1 = 0,797$

$X2 = \frac{121,890 - 63,810}{128,530 - 63,810}(0,8) + 0,1 = 0,818$

$X2 = \frac{77,770 - 63,810}{128,530 - 63,810}(0,8) + 0,1 = 0,273$

$X3 = \frac{131,350 - 96,890}{180,720 - 96,890}(0,8) + 0,1 = 0,429$

$T1 = \frac{140,000 - 100,000}{200,000 - 100,000}(0,8) + 0,1 = 0,420$

B. Determining the Weighting Value

After the data has been normalized, the next step is to determine the initial weight value used the nguyen widrow algorithm to get the initial weight value.

Table 9. Initial Weight Value

V		V_{ij}	
	1	2	3
V1	0,704	0,386	0,641
V2	0,538	0,635	0,635
V3	0,810	0,295	0,114

The next step is to calculate the layer values using the following equation:

$Zin_1 = 0,679 + (0,797 \times 0,704) + (0,900 \times 0,386) + (0,420 \times 0,641 \ = 1,881637$

$Zin_2 = 0,479 + (0,534 \times 0,538) + (0,818 \times 0,635) + (0,650 \times 0,635 \ = 1,70117$

$Zin_3 = 0,158 + (0,100 \times 0,810) + (0,273 \times 0,295) + (0,676 \times 0,11 = 0,39632$

3.2.3 Weighting Value

The next step is to calculate the backpropagation value to find the weight value in the test data. As shown in the following table.

Table 10. Weighting Value

		Δ_{VIJ}		
J	1	2	3	4
1	0,033	-0,042	-0,004	-0,004
2	-0,038	-0,0047	-0,31	-0,061
3	0,036	-0,022	-0,015	-0,029
....	......	....	....	......
12	-0,041	-0,046	-0,018	-0,01

3.2.4 Annual Predictions

After carrying out the calculations for the v_ij weight values, the next step is to carry out calculations up to the 12th training data to be able to predict the production based on actual values, the calculation shown in the following table.

Table 11. Prediction Results

No	2021	Actual	Prediction	Error
1	12783	336830	33175	0,032
2	15800	379000	377824	0,031
3	15625	303160	30815	0,021
4	16529	435259	434646	0,013
5	15452	435250	434646	0,024
…	…	….	….	….
11	1807	42761	42802	0,002
12	16529	435259	434646	0,013

The actual value is close to the predicted value, which means that the predicted results have reached the predetermined target.

$$MSE = \frac{(336830 - 33175)^2}{2} = 0{,}032$$
$$MSE = \frac{(379000 - 377824)^2}{2} = 0{,}031$$
$$MSE = \frac{(303160 - 30815)^2}{2} = 0{,}021$$

Where the actual data approaches the predicted value determined by the mean square error value close to 0,001. The following is a visualization of the results using the backpropagation neural network method.

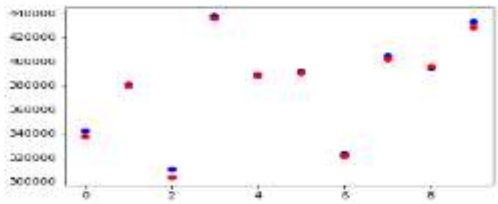


Figure 4 BPNN's result visualization

Where the red color is the result of the prediction and the blue color is the actual data.

3.2.5 Comparing with the Fuzzy Time Series Method

This research uses palm oil production data from PKS Cot Girek from 2019 to 2021.

Table 12. Prediction Result

Actual Data	Prediction Result
131.350	120.712
154.510	130.321
156.250	134.876
158.230	146.859
127.830	159.885
180.720	160.823
96.890	170.055
155.320	194.770
165.290	205.919
143.700	206.645
158.000	207.575

Next, a comparison process was carried out, where the most accurate prediction method in this reseach was the backpropagation neural network method, compared to the fuzzy time series. Where the predicted value using the backpropagation method is close to the predicted value with a mean square error (mse) close to 0.001% compared to the fuzzy times series where the predicted data is still far from the actual data. The following is the accuracy of the four methods which can be seen in the following diagram.

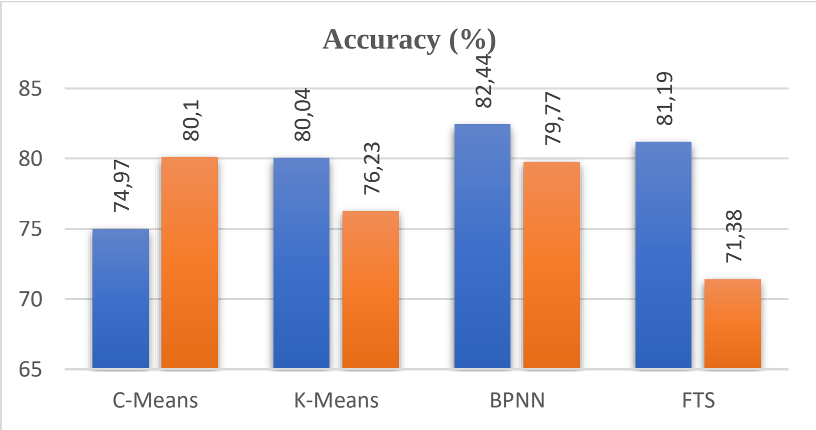


Figure 5.Accuracy Values Between All Methods

The accuracy results by comparing the four methods are the C-Means method is better than K-Menas for clustering because the accuracy value is higher than the K-Means. As for the prediction results, the backpropagation method is better in predicting in terms of the accuracy and also the predicted value is close to the actual value.

4. CONCLUSIONS

The results of the comparative analysis of the C-means and k-means grouping methods. Variable data used Area (Ha), Main and Palm Bunches. The data used was 234 consisting of 134 training data and 100 dataset. The accuracy test for C-Means model values is more accurate with 100 data sets, iterating on C-means is 2 iterations with values C1 (55%), C2 (20%), C3 (35%). The value of the first iteration is 1.87, and the value of the second iteration is 38.7. While the results of k-means iteration up to 4 times. The last iteration value is 4.165. Whereas for prediction with a period of 3 years, the results of prediction of production with backpropagation are 89.77%. while the FTS accuracy rate is 81.38%. The BPNN error value on data is 0.048 with a difference of 0.03%. Production prediction is carried out by using the backpropagation neural network method. The prediction rate with the BPNN method is better than the FTS method, where the accuracy of the prediction with the backpropagation method is compared to the fuzzy time series with an accuracy value of 79.77% BPNN and 71.38% FTS. Where is the fuzzy time series method, the actual value and predictive value are still far away and the error is still high compared to the backpropagation method. For further research, interval optimization on fuzzy time series is suggested to combine with particle swarm optimization to determine the best interval to get the optimum solution for the estimated cost (MSE) value. Some suggestions for this research and further research are the data provided by the company must be better and more complex to be used in grouping each region.

ACKNOWLEDGEMENTS

Thanks to Cot Girek PKS Company and PT Perkebunan Nusantara 1 (Persero) for providing data to support this research. Thank you to all lecturers of information systems, Faculty of Engineering, Universitas Malikussaleh, who have motivated this research to be completed.

REFERENCES

[1] "Analisis Ekspor Minyak Kelapa Sawit (CPO) Indonesia | Abidin | Jurnal Aplikasi Manajemen."

[2] "Kementan : Produksi Minyak Sawit Diperkirakan," p. 2021, 2021.

[3] S. Bari, T. H. Lim, and C. W. Yu, "Effects of preheating of crude palm oil (CPO) on injection system, performance and emission of a diesel engine," *Renew. Energy*, vol. 27, no. 3, pp. 339–351, Nov. 2002, doi: 10.1016/S0960-1481(02)00010-1.

[4] A. Nurkholis and I. S. Sitanggang, "Optimalisasi model prediksi kesesuaian lahan kelapa sawit menggunakan algoritme pohon keputusan spasial," *J. Teknol. dan Sist. Komput.*, vol. 8, no. 3, pp. 192–200, Jul. 2020, doi: 10.14710/JTSISKOM.2020.13657.

[5] D. F. Pasaribu, I. S. Damanik, E. Irawan, Suhada, and H. S. Tambunan, "Memanfaatkan Algoritma K-Means Dalam Memetakan Potensi Hasil Produksi Kelapa Sawit PTPN IV Marihat," *BIOS J. Teknol. Inf. dan Rekayasa Komput.*, vol. 2, no. 1, pp. 11–20, 2021, doi: 10.37148/bios.v2i1.17.

[6] L. D. Yulianto, A. Triayudi, and I. D. Sholihati, "Implementation Educational Data Mining For Analysis of Student Performance Prediction with Comparison of K-Nearest Neighbor Data Mining Method and Decision Tree C4.5," *J. Mantik*, vol. 4, no. 1, pp. 441–451, May 2020.

[7] J. Sains dan Teknologi, P. Satya Saputra, G. Rasben Dantes, and I. Gede Aris Gunadi, "Perbandingan Algoritma Fuzzy C-Means Dan Algoritma Naive Bayes Dalam Menentukan Keluarga Penerima Manfaat (Kpm) Berdasarkan Status Sosial Ekonomi (Sse) Terendah," *JST (Jurnal Sains dan Teknol.*, vol. 10, no. 1, pp. 1–8, Mar. 2021, doi: 10.23887/JSTUNDIKSHA.V10I1.23340.

[8] N. Agustina and P. Prihandoko, "Perbandingan Algoritma K-Means dengan Fuzzy C-Means Untuk Clustering Tingkat Kedisiplinan Kinerja Karyawan," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 2, no. 3, pp. 621–626, 2018, doi: 10.29207/resti.v2i3.492.

[9] A. Latifa, R. Putri, and N. Dwidayati, "Analisa perbandingan k-means dan fuzzy c-means dalam pengelompokan daerah penyebaran COVID-19 Indonesia," *Unnes J. Math.*, vol. 10, no. 2, pp. 50–55, Dec. 2021, doi: 10.15294/UJM.V10I2.50433.

[10] M. C. Bagdatlı and O. Arslan, "Classification And Mapping Of Land Use And Some Soil Properties In Kirşehir Province, Turkey," *Int. J. Eng. Technol. Manag. Res.*, vol. 8, no. 8, pp. 81–93, Sep. 2021, doi: 10.29121/IJETMR.V8.I8.2021.1022.

[11] G. F. Fan, Y. H. Guo, J. M. Zheng, and W. C. Hong, "A generalized regression model based on hybrid empirical mode decomposition and support vector regression with back-propagation neural network for mid-short-term load forecasting," *J. Forecast.*, vol. 39, no. 5, pp. 737–756, Aug. 2020, doi: 10.1002/FOR.2655.

[12] W. D. H. Hendra Effendi, Ahmad Syahrial, Sefran Prayoga, "Penerapan Metode K-Means Clustering untuk Pengelompokan Lahan Sawit Produktif pada PT Kasih Agro Mandiri," *Teknomatika*, vol. 11, no. 02, pp. 117–126, 2021.

[13] H. Aini, H. Haviluddin, E. Budiman, M. Wati, and N. Puspitasari, "Prediksi Produksi Minyak Kelapa Sawit Menggunakan Metode Backpropagation Neural Network," *Sains, Apl. Komputasi dan Teknol. Inf.*, vol. 1, no. 1, p. 24, 2019, doi: 10.30872/jsakti.v1i1.2261.

[14] A. Octaviani and P. Dewi, "Big Data di Perpustakaan dengan Memanfaatkan Data Mining," *Anuva J. Kaji. Budaya, Perpustakaan, dan Inf.*, vol. 4, no. 2, pp. 223–230, Jun. 2020, doi: 10.14710/ANUVA.4.2.223-230.

[15] S. B. Faradilla, "Komparasi Analisis K-Medoids Clustering dan Hierarchical Clustering (Studi Kasus: Data Kriminalitas di Indonesia Tahun 2020)," Apr. 2022.

[16] A. Nofiar, S. Defit, and Sumijan, "Penentuan Mutu Kelapa Sawit Menggunakan Metode

K-Means Clustering," *Jurnal KomtekInfo*, vol. 5, no. 3. pp. 1–9, 2019. doi: 10.35134/komtekinfo.v5i3.26.

[17] F. W. Nugraha, S. Fauziati, and A. E. Permanasari, "SISTEM PENDUKUNG KEPUTUSAN PEMILIHAN VARIETAS KELAPA SAWIT DENGAN METODE FUZZY C-MEANS," *Pros. SENIATI*, vol. 3, no. 1, p. B25. 1-8, 2017, doi: 10.36040/SENIATI.V3I1.1761.

[18] A. R. Lubis, S. Prayudani, Y. Fatmi, Al-Khowarizmi, Julham, and Y. Y. Lase, "Measurement of Detection Rate Accuracy in Forecasting Crude Palm Oil Production using Fuzzy Time Series," *2021 2nd Int. Conf. Innov. Creat. Inf. Technol. ICITech 2021*, pp. 20–24, Sep. 2021, doi: 10.1109/ICITECH50181.2021.9590172.

[19] Rasna, I. W. Sudarsana, and D. Lusiyanti, "Forecasting Of Crude Palm Oil By Using Fuzzy Time Series Method (Study Case : PT. Buana Mudantara Plantation)," *Param. J. Stat.*, vol. 1, no. 1, pp. 31–40, Jan. 2021, doi: 10.22487/27765660.2021.V1.I1.15442.

[20] M. Nishom, "Perbandingan Akurasi Euclidean Distance, Minkowski Distance, dan Manhattan Distance pada Algoritma K-Means Clustering berbasis Chi-Square," *J. Inform. J. Pengemb. IT*, vol. 4, no. 1, pp. 20–24, 2019, doi: 10.30591/jpit.v4i1.1253.

[21] D. L. Rahakbauw, V. Y. I. Ilwaru, and M. H. Hahury, "Implementasi Fuzzy C-Means Clustering Dalam Implementation Of Fuzzy C-Means Clustering In," *J. Ilmu Mat. dan Terap.*, vol. 11, pp. 1–12, 2017.

[22] Y. H. Syahputra and J. Hutagalung, "Superior Class to Improve Student Achievement Using the K-Means Algorithm," *Sink. J. dan Penelit. Tek. Inform.*, vol. 7, no. 3, pp. 891–899, Jul. 2022, doi: 10.33395/SINKRON.V7I3.11458.

[23] N. L. R. Amalia, A. A. Supianto, N. Y. Setiawan, V. Zilvan, A. R. Yuliani, and A. Ramdan, "Student Academic Mark Clustering Analysis and Usability Scoring on Dashboard Development Using K-Means Algorithm and System Usability Scale," *J. Ilmu Komput. dan Inf.*, vol. 14, no. 2, pp. 137–143, Jul. 2021, doi: 10.21609/JIKI.V14I2.980.

[24] H. Effendi, A. Syahrial, S. Prayoga, and W. D. Hidayat, "Penerapan Metode K-Means Clustering Untuk Pengelompokan Lahan Sawit Produktif Pada PT Kasih Agro Mandiri," *Teknomatika*, vol. 11, no. 02, pp. 117–126, Oct. 2021.

[25] N. Nurmila, A. Sugiharto, and E. A. Sarwoko, "Algoritma Back Propagation Neural Network Untuk Pengenalan Pola Karakter Huruf Jawa," *J. Masy. Inform.*, vol. 1, no. 1, pp. 1–10, 2010, doi: 10.14710/jmasif.1.1.

[26] B. Endaryati and R. Kurniawan, "Komparasi Metode Peramalan Automatic Clustering Technique and Fuzzy Logical Relationships Dengan Single Exponential Smoothing," *Media Stat.*, vol. 8, no. 2, pp. 93–101, 2015, doi: 10.14710/medstat.8.2.93-101.

[27] F. A. Hizham, Y. Nurdiansyah, and D. M. Firmansyah, "Implementasi metode Backpropagation Neural Network (BNN) dalam sistem klasifikasi ketepatan waktu kelulusan mahasiswa," *Berk. Sainstek*, vol. 6, no. 2, pp. 97–105, 2018.

[28] M. Nor Hayati and D. Sri Wahyuningsih, "Peramalan Menggunakan Metode Fuzzy Time Series Cheng Forecasting Using Fuzzy Time Series Cheng Method," *J. EKSPONENSIAL*, vol. 8, no. 1, pp. 51–56, 2017.