Available online to ejournal.pnc.ac.id

Journal of Innovation Information Technology and Application

Journal Page is available to https://ejournal.pnc.ac.id/index.php/jinita

Clustering Productive Palm Land using the K- Means Clustering Algorithm

Geofanny Widianto Sihite[1], Eka Prasetyaningrum[2]*

[1,2] *Program Studi Sistem Informasi, Universitas Darwan Ali, Sampit, Indonesia*

email:[1] geofannysihite@gmail.com, [2] eka.tya94@gmail.com

ARTICLE INFO

Article history:
Received 14 August 2023
Revised 21 September 2023
Accepted 21 December 2023
Available online 30 December 2023

Keywords:
Oil Palm
K-Means Algorithm
Clustering
Oil Palm Plantations
Segmentation.

IEEE style in citing this article:
G. Widianto Sihite and E. Prasetyaningrum, "Clustering Productive Palm Land using the K- Means Clustering Algorithm," Journal of Innovation Information Technology and Application (JINITA), vol. 5, no. 2, pp. 141–153, Dec. 2023.

ABSTRACT

Indonesia is a country with a tropical climate that has many oil palm plantations. CV. Alkema Deo is one of the companies that manage oil palm plantations in Sampit City, East Kotawaringin Regency, Central Kalimantan. In this study, a qualitative approach was applied using a descriptive research pattern. In qualitative research, data is obtained from sources using various data collection techniques. Research using qualitative methods emphasizes the analysis of thought processes related to the dynamics of the relationship between observed phenomena, and always uses scientific logic. Based on the results of research for authors on a CV. Alkema Deo, the use of Excel in companies is quite good at processing data, but on a CV. Alkema Deo does not yet have land groupings based on productivity levels, so it is difficult to see the level achieved in 6 months based on the set target, and daily production control in terms of area and block area. Data obtained from CV. Alkema Deo is grouped based on area, block, and productivity. Application of data mining for grouping productive oil palm land on a CV. Alkema Deo with 4 variables, namely: land area, length, average production yield, percentage of achievement using the K-Means Algorithm to produce three clusters, namely 8 blocks or 50% including the high productive group (C2), 1 block or 6% blocks including the medium productive plantation group (C1), and 7 blocks or 44% including the small productive plantation group (C0).

1. INTRODUCTION

Indonesia is a country with a tropical climate that has many plantations, one of which is an oil palm plantation. Oil palm plantations in Indonesia are managed by both the government and private companies[1]. According to the Central Statistics Agency (BPS), the area of oil palm plantations in Indonesia is expected to reach 14.62 million hectares in 2021. This is a slight increase of 0.24% compared to the previous year's 14.59 million hectares. The data shows that the area of Indonesian oil palm plantations has been steadily increasing, particularly in the last decade, with a jump of 60.11% between 2011 and 2021[2]. Did you know that East Kotawaringin is the district with the largest area of oil palm plantations in Indonesia? According to data from the Ministry of Agriculture, the total area of oil palm land in various districts of Central Kalimantan Province is approximately 425,000 hectares (Ha). This amount comprises 406,000 hectares of national and foreign private plantation land, as well as 19,000 hectares of community plantations. Interestingly, the total area of oil palm in Central Kalimantan is 1.48 million hectares, which is the second largest on the island of Kalimantan after West Kalimantan with an area of 1.5 million hectares[3].

Alkema Deo is a company that manages oil palm plantations in Sampit City, located in the East Kotawaringin Regency of Central Kalimantan. Established in 2016, the company has two plantation sites situated on Jl. General Sudirman Km. 18, East Kotawaringin and Seibabi Village, Telawang District, East Kotawaringin. Alkema Deo operates two divisions of oil palm land, which are further divided into 16 blocks, covering a total area of approximately 33 hectares. Each block has an average size of 3-2 hectares and contains around 272 highly efficient palm trees. Each block can produce roughly 3.2 tons of palm fruit per hectare each month. The oil palm land belongs to CV. Alkema

*) Corresponding Author : geofannysihite@gmail.com

Deo is organized by area, block, and productivity. This system of grouping is designed to enable daily production control and land management, to monitor yields effectively. This is critical to ensuring that palm oil production runs smoothly, while also maintaining the land and detecting any errors or declining production in each block. Cluster analysis is a technique used to group objects that share similar characteristics and properties. The objective is to create homogeneous groups with members that share common traits[4].

The k-Means algorithm is a process that helps to group data sets into several clusters. It is a straightforward and adaptable algorithm that is widely used. K-Means clustering is a significant part of data mining. The clustering process is dependent on the data and conclusions are drawn from it. Clustering in data mining is highly beneficial for identifying distribution patterns and analyzing data[5]. Data mining is the process of uncovering knowledge from databases. It involves statistical analysis, mathematics, artificial intelligence, and machine learning to extract useful information and related insights from massive databases[6].

According to the given information, information about productive oil palm land is in the CV. Alkema Deo is needed to utilize the K-Means clustering algorithm method. Pulungan, I. M., Saifullah, S., Fauzan, M., and Windarto, A. P. researched determining the most productive oil palm blocks using the K-Means clustering algorithm. They grouped oil palm plant blocks based on data[7]. Presents other previous studies that used different criteria in applying the K-Means clustering algorithm. Research by Hajar, S., Novany, AA, Windarto, AP, Wanto, A., & Irawan, E (2020) entitled Application of K-Means Clustering on palm oil exports by destination country. The research results show that palm oil exports can be completed using the K-Means Clustering Algorithm Method [8]. Research by Nofiar, Andri, dan Sarjon Defit. Tambunan, HS (2018) entitled Determining Palm Oil Quality Using the K-Means Clustering Method. The research results of the k-means clustering method can be used to process data using data mining concepts in grouping data based on criteria [9]. Research by Nuraisana, Nuraisana (2019) entitled Clustering analysis to determine the level of oil palm potential based on crop area using the K-means algorithm at the North Sumatra statistical agency. The research results are widely known that the use of the K-Means algorithm can assess the prospect of oil palm plantation capacity according to the area of cultivated land in North Sumatra province[10]. Research by Febiola, Yessica Inggir, Imam Cholissodin, dan Candra Dewi (2019) entitled Forecasting Palm Oil Yields Using the Multifactor High Order Fuzzy Time Series Method Optimized with K-Means Clustering (Case Study: PT. Sandabi Indah Lestari, Bengkulu City). Research results It is known that this research used several factors consisting of monthly oil palm yield, land area, plant age, and amount of palm oil staples [11]. Research by Sari, KAMI, Muslimin, M., Franz, A., & Sugiartawan, P. (2022) entitled Detection of Maturity Levels of Fresh Palm Fruit Bunch using the K-Means Algorithm. The results of the K-Means clustering research are a method of grouping objects based on their proximity to the average value of the cluster center for a certain number of clusters, k [12]. Research by Pratama, Faiq Husain, Agung Triayudi, dan Eri Mardiani (2022) entitled K-medoids and k-means data mining for the classification of palm oil production potential in Indonesia. This is because there are similarities in the characteristics of plantations based on similarities in area, production, and productivity [13]. Based on previous research, this research is to classify productive oil palm land in CV. Alkema Deo with 4 variables, namely land area, land area, average production yield, and percentage of achievement using data mining methods.

Therefore, this research aims to analyze and identify groupings of highly productive, medium productive, and small productive oil palm lands using the K-Means clustering algorithm method. Apart from that, it also takes into account the variables of land area, number of rows, average production yield, and percentage of achievement of the specified targets as a basis for grouping. Based on this objective, the benefit of this research is that companies can find out the grouping of oil palm land based on the productive level so that it can help in land management and make decisions that are expected to increase oil palm production results.

2. RESEARCH METHOD

This study applied a qualitative approach using a descriptive research pattern. The secondary data used in this study is plantation production data based on outside areas in CV. Alkema Deo from January to June 2022. The object of research in this study is the area of the CV. Alkema Deo is located at Jendral Sudirman Km.18 and in the village of Seibabi, Telawang District, East Kotawaringin Regency. The area to be studied is 33 hectares (ha) consisting of 2 divisions totaling 16 blocks of productive oil palm land. Based on the research object, it can be determined that the population is all the oil palm land in CV. Alchemy God. Then for the research sample are productive oil palm lands based on division, block, area/land, empty fruit bunch, average yield, and yielding percentage.

2.1. Data Collection Methods

Production data for 2022 is then processed by data mining with the following steps: data mining with the following steps; Data cleaning (Data Cleaning); Data Integration (Data Integration); Data Selection (Data Selection); Transformation Data (Data Transformation); and Presentation of Knowledge obtained as a result of data mining processing. Calculations were performed using Excel software [14].

To carry out the grouping of oil palm land in CV. Alkema Deo uses secondary data as mentioned, with the following steps in data mining processing:

1. Data Cleaning:
    a. Identify and treat missing or blank values in data such as area, empty fruit bunch, average yield, and percentage gain.
    b. Identification and handling of invalid or inconsistent data, for example, data that exceeds logical limits or differs from the specified format.
    c. Eliminate data duplication if any.
2. Data Integration:
    a. Combining data from various relevant sources, such as data on area, empty fruit bunch, average yield, and percentage of achievement, into one integrated dataset.
    b. Ensuring uniformity and consistency of data formats between columns.
3. Data Selection:
    a. Select the most relevant and significant variables or attributes to group oil palm lands.
    b. Remove variables that are irrelevant or have little effect on grouping.
4. Transformation Data (Data Transformation):
    a. Perform data transformations when necessary, such as data normalization or scaling, to ensure uniformity and consistency of data in the dataset.
    b. Apply appropriate methods or techniques to change data formats or representations where necessary.
5. Knowledge Presentation:
    a. Use appropriate data mining processing techniques, such as clustering algorithms, to classify oil palm areas based on existing data.
    b. Analyze grouping results and present them in informative presentations, such as tables, graphs, or data visualizations.

All steps of data mining processing can be done using Excel software. You can take advantage of various features and functions of Excel to clean, integrate, select, transform data, and present the knowledge obtained through processing the data mining.

2.2 Data Processing Process

Production data for 2022 will then be processed using the data mining method with the following stages: data cleaning; data integration; data selection; data transformation; data mining; and knowledge representation obtained from the results of the data mining processing[15]. It is a process for each stage of data processing. The computational process is carried out using Excel and RapidMiner software to simplify data processing.

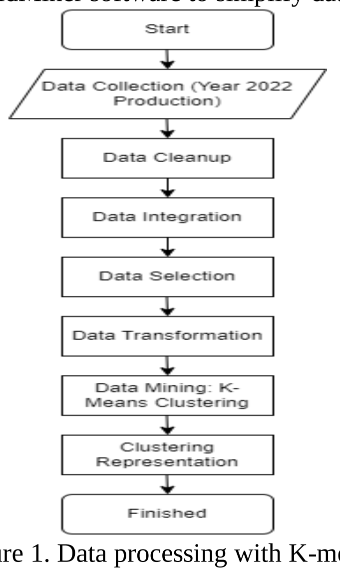

Figure 1. Data processing with K-means

DOI: doi.org/ 10.35970/ jinita.v5i2.2051

Data mining with the K-Means Clustering technique will be carried out after the data is transformed. The grouping of the data is carried out using the stages shown in Figure 1. Determining the centroid can use the Euclidean formula as follows [15]:

$$\sqrt{\sum_{i=1}^{n}(Xi - Yj)^2}$$

Description (1)
d = distance between a point and the center point,
Xi = the value of the ith variable in up to X,
Yj = the value of the jth variable from the clustering center Y

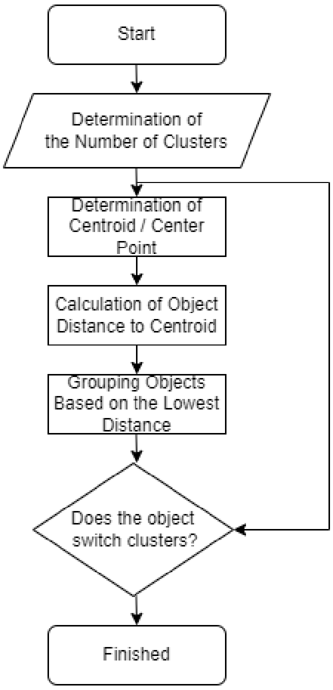

Figure 2 Segmentation Process Flow with K-Means

Segmentation with the K-Means Clustering technique in data mining involves several stages. The procedure for each stage of this technique can be carried out as shown in Figure 2. The calculation results from data processing using this technique will produce a segmentation of oil palm plantations in the CV. Alkema Deo is divided into several groups based on the same results in the 2022 production data[16]. The analysis step is the last in the plantation segmentation process. After the data is processed using the K-Means clustering technique, the data will be further analyzed to find out the details of the oil palm plantation segments in various areas of CV. Alkema Deo is based on the similarity of area characteristics (Ha), the number of shoots produced, the production yield of the specified target (%), and the average production yield per hectare (kg) [16].

3. RESULTS AND DISCUSSION

3.1. Research result

Based on the results of research for writers on CVs. Alkema Deo, the use of Excel in companies is quite good at processing data, but on a CV. Alkema Deo does not yet have land groupings based on productivity levels, so it is difficult to see the level achieved in 6 months based on the set target, and daily production control in terms of area and block area.

3.2. Discussion

3.2.1. Data and Information Collection

Data obtained from the CV. Alkema Deo is grouped by area, block, and productivity. This grouping aims to provide daily production control, and land control to monitor yields, which play an important role in ensuring that palm oil production continues to run well, maintaining the land, and seeing if there are any errors or decreased production in each block according to production data in 2022. Oil palm plantation data to see the potential generated if there are 16 data shown in Table 2. Production data for 2022 consists of block data, Op, area (Ha), principal amount, empty fruit bunch, production, and Production Distribution.

Table 2. Production data for 2022

No	Block data	OP	Area (Ha)	Principal amount	Empty fruit bunch	Production	PRODUCTION SPREAD (KG)					
							January	February	March	April	May	June
1	1.A	2014	3	408	810	3260	3875	3469	3550	3690	3550	3880
2	1.B	2014	2	272	569	2000	2780	1980	2100	2371	2500	2790
3	1.C	2014	2	272	545	2140	2576	2081	2370	2433	2400	2450
4	1.D	2014	2	272	480	1200	1600	2000	1850	1750	1760	2000
5	1.E	2014	2	272	465	1180	2001	1750	2100	2000	1900	2100
6	1.F	2014	2	272	486	1350	2280	2200	2350	2100	2150	2030
7	1.G	2014	2	272	520	1900	2150	2200	2375	2250	2325	2100
8	1.H	2014	2	272	578	2150	2790	2700	2825	2750	2600	2850
9	1.I	2014	2	272	563	1890	2518	2760	2340	2516	2625	2400
10	2.A	2015	2	270	421	1500	1890	1900	2000	2150	2110	2125
11	2.B	2015	2	270	480	1770	2100	2050	2150	2162	2200	2175
12	2.C	2015	2	270	428	1270	1970	1900	1950	2000	1910	2050
13	2.D	2015	2	270	450	1540	2230	1915	2000	2150	2100	2150
14	2.E	2015	2	270	317	1350	1850	1600	1650	1788	1900	1850
15	2.F	2015	2	270	459	1700	2000	1915	1950	2100	2150	2000
16	2.G	2015	2	270	280	1100	1190	1090	1100	1050	1080	1050

Data source: Processed 2023 data

3.2.2. Data Cleanup

The data contained in Table 2 is production data for 2022 which the author obtained from CV. Alkema Deo has no data duplication or inconsistent data, so data cleaning is not necessary.

3.2.3. Data Integration

The data that the author will use in the data mining process consists of one data source, namely production data for 2022. Production data for 2022 can be seen in Table 2, and integration data can be seen in Table 3.

Table 3. Data Integration

No	Division	Year of Planting	Block	Area	Principal	Number Empty fruit bunch	Production	Target
1	1	2014	1.A	3	408	810	3260	3875
2	1	2014	1.B	2	272	569	2000	2780
3	1	2014	1.C	2	272	545	2140	2576
4	1	2014	1.D	2	272	480	1200	1600
5	1	2014	1.E	2	272	465	1180	2001
6	1	2014	1.F	2	272	486	1350	2280
7	1	2014	1.G	2	272	520	1900	2150
8	1	2014	1.H	2	272	578	2150	2790
9	1	2014	1.I	2	272	563	1890	2518
10	2	2015	2.A	2	270	421	1500	1890
11	2	2015	2.B	2	270	480	1770	2100
12	2	2015	2.C	2	270	428	1270	1970
13	2	2015	2.D	2	270	450	1540	2230
14	2	2015	2.E	2	270	317	1350	1850
15	2	2015	2.F	2	270	459	1700	2000
16	2	2015	2.G	2	270	280	1100	1190

Data source: Processed 2023 data

DOI: doi.org/ 10.35970/ jinita.v5i2.2051

3.2.4. Data Selection

The data cleaning that has been done is then re-selected and the right data is selected for the process of grouping CV. Alkema Deo productive oil palm land. Data selection can be seen in Table 4.

Table 4. Data Selection

Division	Block	Area	Empty Fruit Bunch	Production	Target
1	1.A	3	810	3260	3875
1	1.B	2	569	2000	2780
1	1.C	2	545	2140	2576
1	1.D	2	480	1200	1600
1	1.E	2	465	1180	2001
1	1.F	2	486	1350	2280
1	1.G	2	520	1900	2150
1	1.H	2	578	2150	2790
1	1.I	2	563	1890	2518
2	2.A	2	421	1500	1890
2	2.B	2	480	1770	2100
2	2.C	2	428	1270	1970
2	2.D	2	450	1540	2230
2	2.E	2	317	1350	1850
2	2.F	2	459	1700	2000
2	2.G	2	280	1100	1190

Data source: Processed 2023 data

3.2.5. Data Transformation

Data transformation resulted in 16 datasets which would be processed using the K-Means Clustering technique. Before executing using this technique, each variable is assigned a special attribute to facilitate the processing of the selected data, which is converted to a format that suits your needs in the form of:

a. Average production yield = $\frac{Production}{Area}$

b. Achievement percentage = $\frac{Production}{Target} * 100$

produce the right data when used in the data mining process k-means Table 5 data transformation.

Table 5. Data transformation

No.	Division	Block	Area	Empty Fruit Bunch	Production	Average Production Yield	Target	Achievement Percentage
1	1	1.A	3	810	3260	1086,66	3875	84,12
2	1	1.B	2	569	2000	1000	2780	71,94
3	1	1.C	2	545	2140	1070	2576	83,07
4	1	1.D	2	480	1200	600	1600	75
5	1	1.E	2	465	1180	590	2001	58,97
6	1	1.F	2	486	1350	675	2280	59,21
7	1	1.G	2	520	1900	950	2150	88,37
8	1	1.H	2	578	2150	1075	2790	77,06
9	1	1.I	2	563	1890	945	2518	75,05
10	2	2.A	2	421	1500	750	1890	79,36
11	2	2.B	2	480	1770	885	2100	84,28
12	2	2.C	2	428	1270	635	1970	64,46
13	2	2.D	2	450	1540	770	2230	69,05
14	2	2.E	2	317	1350	675	1850	72,97
15	2	2.F	2	459	1700	850	2000	85
16	2	2.G	2	280	1100	550	1190	92,43

Data source: Processed 2023 data

3.2.6. Clustering K-means

Based on data that has been processed in the early stages of data mining, land area, number of stems, average yield and percentage of achievement are the variables that will be used in clustering calculations. The following are the steps for calculating the K-Means algorithm using a sample of January 2022 production data, table 6 January 2022 Production Data.

Table 6. January 2022 Production Data.

No.	Division	Block	Area	Empty Fruit Bunch	Production	Average Production Yield	Target	Achievement Percentage
1	1	1.A	3	810	3260	1086,66	3875	84,12
2	1	1.B	2	569	2000	1000	2780	71,94
3	1	1.C	2	545	2140	1070	2576	83,07
4	1	1.D	2	480	1200	600	1600	75
5	1	1.E	2	465	1180	590	2001	58,97
6	1	1.F	2	486	1350	675	2280	59,21
7	1	1.G	2	520	1900	950	2150	88,37
8	1	1.H	2	578	2150	1075	2790	77,06
9	1	1.I	2	563	1890	945	2518	75,05
10	2	2.A	2	421	1500	750	1890	79,36
11	2	2.B	2	480	1770	885	2100	84,28
12	2	2.C	2	428	1270	635	1970	64,46
13	2	2.D	2	450	1540	770	2230	69,05
14	2	2.E	2	317	1350	675	1850	72,97
15	2	2.F	2	459	1700	850	2000	85
16	2	2.G	2	280	1100	550	1190	92,43

Data source: Processed 2023 data

Steps for Completion of K-Means:

1. The number of clusters formed is 3 clusters. Where cluster 2 (C2) is highly productive, cluster 1 (C1) is productive and cluster 0 (C0) is low productive.
2. In testing as the center of the cluster starting point (centroid) the author uses values from production data for January 2022 by determining the maximum value for C0, the average value for C1, and the minimum value for C2. Following are the results of the initial centroid data for iteration 1, as shown in Table 7:

Table 7. Initial Centroid

Cluster/ Atribut	Area	Empty Fruit Bunch	Average Production Yield	Achievement Percentage
C0	2	578	1075	77,06
C1	3	810	1086,66	84,12
C2	2	317	675	72,97
C0	2	578	1075	77,06

Data source: Processed 2023 data

In Table 7 above, the initial centroid is determined. The optimal number of clusters is determined by looking at the maximum average value of the silhouette S(i). The number of K-means clusters is an estimate of the variable that maximizes the average value of the silhouette validity index S(i). For cluster 0(C0) Area = 2, empty fruit bunch= 578, Average yield = 1075, Percentage of Achievement = 77.06. cluster 1(C1) Area = 3, empty fruit bunch= 810, Average yield = 1086.66, Percentage of Achievement = 84.12. Cluster 2(C2) Area = 2, empty fruit bunch = 317, Average production yield = 675, Achievement Percentage = 72.97.

3. Calculation of iteration 1 by calculating the shortest distance using the Euclidean Distance formula:

1AC0 = $\sqrt{(3-2)^2 + (810-578)^2 + (1086{,}66-1075)^2 + (84{,}12-77{,}06)^2}$ 1AC0 = 232,4022

From the results of the calculation above, it is obtained that the distance between production data and block 1A cluster zero is 232.4022.

1AC1= $\sqrt{(3-3)^2 + (810-810)^2 + (1086{,}66-1086{,}66)^2 + (84{,}12-84{,}12)^2}$ 1AC1 = 0

From the results of the calculation above, it is obtained that the distance between production data and the first block 1A cluster is 0.

1AC2 = $\sqrt{(3-2)^2 + (810-317)^2 + (1086{,}66-675)^2 + (84{,}12-72{,}97)^2}$ 1AC2 = 642,3693

From the results of the calculation above, the result shows that the distance between the production data and the second block 1A cluster is 642.3693.

4. The results of the iteration 1 calculation are seen in Table 8 below:

Table 8. Grouping in Iteration 1

No.	Division	Block	C0	C1	C2	Closest Distance
1	1	1.A	232,4022	0	642,3693	0
2	1	1.B	75,71139	256,3987	411,2543	75,71139
3	1	1.C	33,91342	265,5271	456,1919	33,91342
4	1	1.D	485,0085	588,0664	179,4383	179,4383
5	1	1.E	498,3184	605,2517	171,2454	171,2454
6	1	1.F	410,8316	524,463	169,5592	169,5592
7	1	1.G	138,2639	320,6166	342,1566	138,2639
8	1	1.H	0	232,4022	477,6377	0
9	1	1.I	130,878	284,8856	365,2675	130,878
10	2	2.A	360,9422	514,4751	128,3816	128,3816
11	2	2.B	213,9068	386,7399	266,0769	213,9068
12	2	2.C	465,0363	591,8685	118,2938	118,2938
13	2	2.D	330,8673	479,6891	163,4912	163,4912
14	2	2.E	477,6377	642,3693	0	0
15	2	2.F	254,6548	423,3329	225,685	225,685
16	2	2.G	603,8752	754,3037	131,8055	131,8055
	Total		6	1	9	

Data source: Processed 2023 data

In Table 8 above, the results of iteration 1 calculations obtained grouping results for C0 = 6 (38%), C1 = 1 (6%), and C2 = 9 (56%) as shown in the table above. The iteration process continues until the number of clusters is equal to the number of previous iterations. When the number of the last cluster is the same as the previous cluster, the K-Means process stops. In other words, if the calculation of the resulting data group changes, repeat each data with a new centroid.

5. In testing this sample, the iteration process is carried out 3 times because the number of members from the 2nd and 3rd iteration calculations is the same, the iteration process is stopped. Following are the results of the new centroid for iteration 2.

Table 9. Centroid iteration 2

Cluster/ Atribut	Area	Empty Fruit Bunch	Average Production Yield	Achievement Percentage
C0	2	542,5	987,5	79,96167
C1	3	810	1086,66	84,12
C2	2	420,6667	677,2222	72,93889

Data source: Processed 2023 data

Table 9 above determines the centroid for iteration 2. The optimal number of clusters is determined by looking at the maximum average value of the silhouette S(i). The number of K-means clusters is an estimate of the variable that maximizes the average value of the silhouette validity index S(i). For cluster 0(C0) Area = 2, Empty fruit bunch = 542.5, Average yield = 987.5, Percentage of Achievement = 79.96167. cluster 1(C1) Area = 3, Empty fruit bunch = 810, Average yield = 1086.66, Percentage of Achievement = 84.12. Cluster 2(C2) Area = 2, Empty fruit bunch = 420.6667, Average production yield = 677.2222, Achievement Percentage = 72.93889. After all, points are calculated in the nearest cluster, then do the calculation again on the centroid of Iteration 2.

6. Calculation of iteration 2 by calculating the shortest distance using the Euclidean Distance formula:

1AC0 = $\sqrt{(3-2)^2 + (810-542{,}5)^2 + (1086{,}66-987{,}5)^2 + (84{,}12-79{,}96167)^2}$ 1AC0 = 285,3196

From the results of the calculation above, it is obtained that the distance between production data and block 1A cluster zero is 285,3196.

1AC1= $\sqrt{(3-3)^2 + (810-810)^2 + (1086{,}66-1086{,}66)^2 + (84{,}12-84{,}12)^2}$ 1AC1 = 0

From the results of the calculation above, it is obtained that the distance between production data and the first block 1A cluster is 0.

1AC2 = $\sqrt{(3-2)^2 + (810-420{,}6667)^2 + (1086{,}66-677{,}2222)^2 + (84{,}12-72{,}93889)^2}$ 1AC2 = 565,1069. From the results of the calculation above, the result shows that the distance between the production data and the second block 1A cluster is 565,1069.

7. The results of iteration 2 calculations can be seen in Table 10 below:

Table 10 Grouping in Iteration 2

No.	Division	Block	C0	C1	C2	Closest Distance
1	1	1.A	285,3196	0	565,1069	0
2	1	1.B	30,3784	256,3987	355,2313	30,3784
3	1	1.C	82,59638	265,5271	412,1114	82,59638
4	1	1.D	392,5393	588,0664	97,40618	97,40618
5	1	1.E	405,5282	605,2517	98,83466	98,83466
6	1	1.F	318,2438	524,463	66,79719	66,79719
7	1	1.G	44,53313	320,6166	290,7111	44,53313
8	1	1.H	94,47179	232,4022	427,7826	94,47179
9	1	1.I	47,44075	284,8856	303,2625	47,44075
10	2	2.A	266,7749	514,4751	73,06125	73,06125
11	2	2.B	120,1297	386,7399	216,3808	120,1297
12	2	2.C	370,9539	591,8685	43,68507	43,68507
13	2	2.D	236,6042	479,6891	97,38215	97,38215
14	2	2.E	385,4288	642,3693	103,6905	103,6905
15	2	2.F	160,9468	423,3329	177,3896	160,9468
16	2	2.G	510,3606	754,3037	190,6633	190,6633
	Total		7	1	8	

Data source: Processed 2023 data

Table 10 above is the result of iteration 2 calculations, the grouping results for C0 = 7 (40%), C1 = 1 (6%), and C2 = 8 (50%) are shown in the table above. The iteration process continues until the number of clusters is equal to the number of previous iterations.

Table 11 Centroid iteration 3

Cluster/ Atribut	Area	Empty fruit bunch	Average yield	Percentage of Achievement
C0	2	530,5714	967,8571	80,68143
C1	3	810	1086,66	84,12
C2	2	415,875	655,625	71,43125

Data source: Processed 2023 data

Table 11 above determines the centroid for iteration 3. The optimal number of clusters is determined by looking at the maximum average value of the silhouette S(i). The number of K-means clusters is an estimate of the variable that maximizes the average value of the silhouette validity index S(i). For cluster 0(C0) Area = 2, empty fruit bunch = 530.5714, Average production yield = 967.8571, Achievement Percentage = 80.68143. cluster 1(C1) Area = 3, empty fruit bunch = 810, Average yield = 1086.66, Percentage of Achievement = 84.12. Cluster 2(C2) Area = 2, empty fruit bunch = 415.875, Average production yield = 655.625, Achievement Percentage = 71.43125. After all, points are calculated in the nearest cluster, then do the calculation again at the centroid of Iteration 3.

8. Calculation of iteration 3 by calculating the shortest distance using the Euclidean Distance formula:

1AC0 = $\sqrt{(3-2)^2 + (810-530{,}5714)^2 + (1086{,}66-967{,}8571)^2 + (84{,}12-80{,}68143)^2}$ 1AC0 = 303,6565
From the results of the calculation above, it is obtained that the distance between production data and block 1A cluster zero is 303,6565

1AC1= $\sqrt{(3-3)^2 + (810-810)^2 + (1086{,}66-1086{,}66)^2 + (84{,}12-84{,}12)^2}$ 1AC1 = 0

From the results of the calculation above, it is obtained that the distance between production data and the first block 1A cluster is 0.

1AC2 = $\sqrt{(3-2)^2 + (810-415{,}875)^2 + (1086{,}66-655{,}625)^2 + (84{,}12-71{,}43125)^2}$ 1AC2 = 584,1983

From the results of the calculation above, the result shows that the distance between the production data and the second block 1A cluster is 584,1983.

9. After getting the centroid update, the next step is to repeat the iteration with the new centroid center by doing the same calculations as iteration 1. Then the iteration 3 cluster results are obtained as in Table 12.

Table 12 Grouping in Iteration 3

No.	Division	Block	C0	C1	C2	Closest Distance
1	1	1.A	303,6565	0	584,1983	0
2	1	1.B	50,85598	256,3987	376,8842	50,85598
3	1	1.C	103,1846	265,5271	434,1836	103,1846
4	1	1.D	371,3605	588,0664	84,96406	84,96406
5	1	1.E	384,1185	605,2517	82,91676	82,91676
6	1	1.F	297,0066	524,463	73,77171	73,77171
7	1	1.G	22,13022	320,6166	312,7069	22,13022
8	1	1.H	117,227	232,4022	449,6572	117,227
9	1	1.I	40,07211	284,8856	324,6487	40,07211
10	2	2.A	243,8634	514,4751	94,84604	94,84604
11	2	2.B	97,13766	386,7399	238,5162	97,13766
12	2	2.C	348,6802	591,8685	24,91996	24,91996
13	2	2.D	213,9497	479,6891	119,381	119,381
14	2	2.E	362,5431	642,3693	100,7672	100,7672
15	2	2.F	137,9544	423,3329	199,5633	137,9544
16	2	2.G	487,3691	754,3037	173,3771	173,3771
	Total		7	1	8	

Data source: Processed 2023 data

After getting the centroid update, the next step is to repeat the iteration with the new centroid center by doing the same calculations as iteration 1. Then the iteration 3 cluster results are obtained as in 13.

Table 13 Clustering Results

No.	Block	*Cluster*
1	1.B	*Cluster* 0
2	1.C	*Cluster* 0
3	1.G	*Cluster* 0
4	1.H	*Cluster* 0
5	1.I	*Cluster* 0
6	2.B	*Cluster* 0
7	2.F	*Cluster* 0
8	1.A	*Cluster* 1
9	1.D	*Cluster* 2
10	1.E	*Cluster* 2
11	1.F	*Cluster* 2
12	2.A	*Cluster* 2
13	2.C	*Cluster* 2
14	2.D	*Cluster* 2
15	2.E	*Cluster* 2
16	2.G	*Cluster* 2

Data source: Processed 2023 data

The data mining used in this segmentation is the k-means clustering technique, which divides oil palm plantations into several segments based on the four variables of area, length, average yield, and percentage of achievement. Segmentation is done using RapidMiner software. The first thing to do is to make the K-means design process by importing the data contained in Table 13. The number of clusters formed is 3 clusters. Where cluster 0 (C0) is low productive, cluster 1 (C1) is productive, and cluster 2 (C2). Determination of the centroid value of the center point is obtained randomly and automatically with the type of measurement used in the form of numerical measures of distance. The centroid value in each cluster is based on area, length, average production yield, and the percentage of achievement used by the centroid table as shown in Figure 3. Centroid data in the centroid table.

Attribute	cluster_0	cluster_1	cluster_2
Luas	2	3	2
Janjang	530.571	810	415.875
Rata- rata hasil produksi	967.857	1086.660	655.625
Persentase Pencapaian	80.681	84.120	71.431

Figure 3. The data on the Centroid Table from Rapidminer

The results of the k-means clustering process from Rapidminer with three segmentation models as shown in Figure 4. The results of the segmentation provide information that the first segment or cluster 0 consists of 7 blocks, the second segment, or cluster 1 consists of 1 block, and the third segment, or cluster 2 consists of 8 blocks.

Cluster Model

```
Cluster 0: 7 items
Cluster 1: 1 items
Cluster 2: 8 items
Total number of items: 16
```

Figure 4. Productive Segmentation Results for January Using K-Means Clustering from Rapidminer

3.2.7. K-Means Clustering Representation

The calculation results show that the content of each element of the oil palm plantation segment is different CV. Alchemy God. Representatives of K-Means groupings can be seen based on Figure 5 Representation of K-Means Clustering In Figure Note that cluster 0 is displayed in green, which is a group of plants that produce low oil palm plantations. Cluster 1 blue means the group with moderate yields of oil palm plantations. Cluster 2, shown in orange, is a group of plantations with high yields of oil palm plantations.

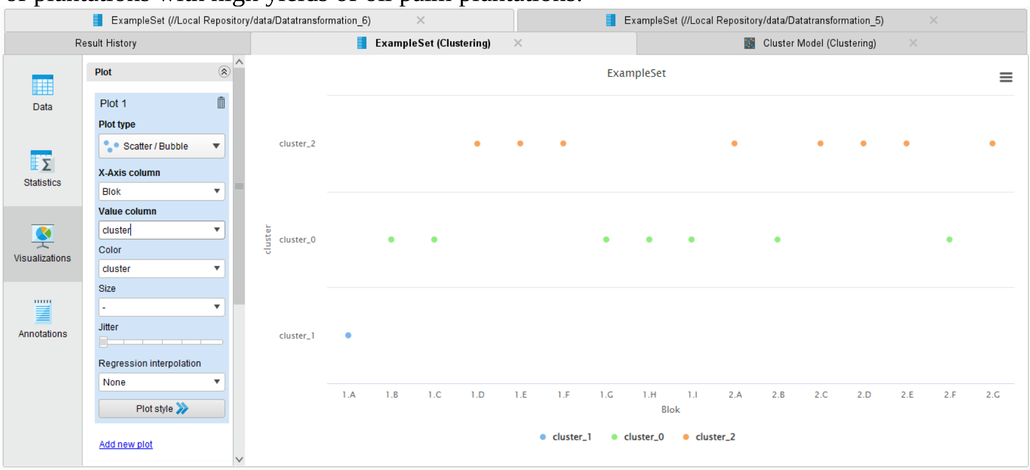

Figure 5 Representation of K-Means Clustering from Rapidminer

3.2.8. Data Analysis

Cluster 0 is a plantation group that has low yield potential for oil palm plantations, consisting of 1B, 1C, 1G, 1H, 1I, 2B, and 2F.

In cluster 0 it can be seen that the total area lies in the range of 2 Ha, length in the range of 530,571 to 459, average production yields in the range of 967,857 Kg/Ha to 850 Kg/Ha, and the percentage of achievement is 80,681 to 71,94.

Cluster 1 is a plantation group that has moderate potential for oil palm plantations, consisting of ID: 1A. In cluster 1 it can be seen that the total area lies in the 3 Ha range, the Length is in the 810 range, the average production yield is in the 1086.660 Kg/Ha range, and the percentage of achievement is 84.120.

Cluster 2 is a plantation group that has high yield potential for oil palm plantations, consisting of ID: 1D, 1E, 1F, 2A, 2C, 2D, 2E, and 2G. In cluster 2 it can be seen that the total area lies in the range of 2 Ha, the Length in the range of 415.875 to 280, and the average yield in the range of 655.625 Kg/Ha to 550 Kg/Ha, and the percentage of achievement is 71.431 to 92.43.

From Table 14 it can be seen that from 100% of the data, 44% of the blocks have the same characteristics in cluster 0, 6% of the blocks have the same characteristics in cluster 1, and 50% of the blocks have the same characteristics in cluster 2.

Table 14. Clustering Results of January Production of CV Alkema in 2022 in Various Blocks

Blok	C0	C1	C2
1.A		√	
1.B	√		
1.C	√		
1.D			√
1.E			√
1.F			√
1.G	√		
1.H	√		
1.I	√		
2.A			√
2.B	√		
2.C			√
2.D			√
2.E			√
2.F	√		
2.G			√

4. CONCLUSION

Based on the results of the evaluation of the clustering formation experiment on the sector. Determination of the value of the initial centroids (central value) of the data is very influential on the success of grouping oil palm land on a CV. Alkema Deo data uses the k-means algorithm. The process of how many times to iterate also influences the suitability of the results of clustering data. By using the K-Means algorithm, 3 clusters, and 8 blocks were obtained, those with high productivity (C2), and medium productivity (C1) were plantation blocks in Seibabi Village, Kec. Telawang, which has low productivity (C0). It is hoped that future research will use comparisons of data processing with other algorithms so that the level of accuracy can be determined even better, and regional data can be added to make it more detailed.

ACKNOWLEDGEMENTS

Praise and thanks to Allah SWT. The researcher expresses immense gratitude to all who supported and assisted in completing this study.

REFERENCES

[1] A. Syahrial, S. Prayoga, and W. D. Hidayat, “Pengelompokan Lahan Sawit Produktif Menggunakan Metode K-Means Clustering Pada PT Kasih Agro Mandiri,” 2021.

[2] S. Sadya, “Perkebunan Kelapa Sawit Indonesia Capai 14,62 Juta Ha pada 2021,” *dataindonesia.id*, 2022. https://dataindonesia.id/agribisnis-kehutanan/detail/perkebunan-kelapa-sawit-indonesia-capai-1462-juta-ha-pada-2021 (accessed Dec. 05, 2022).

[3] Tim Publikasi, “Kotawaringin Timur, Kabupaten dengan Lahan Sawit Terluas di Indonesia,” *https://databoks.katadata.co.id/*, 2019. https://databoks.katadata.co.id/datapublish/2019/10/24/kotawaringin-timur-kabupaten-dengan-lahan-sawit-terluas-di-indonesia (accessed Oct. 24, 2019).

[4] B. Poerwanto and R. Y. Fa’rifah, “Algoritma k-means dalam mengelompokkan kecamatan di tana luwu berdasarkan produktifitas hasil pertanian,” *J. Chem. Inf. Model.*, vol. 53, no. 9, pp. 1689–1699, 2019.

[5] S. Junia Sindy, “Penerapan Metode K-Means Clustering Untuk Menentukan Status Mutu Produksi Buah Kelapa Sawit Pada PT Sawit Asahan Indah (SAI),” *Diss. Univ. Pasir Pengaraian*, 2020.

[6] K. Sawit, D. Ppks, R. E. Marpaung, J. T. Hardinata, and R. A. Nasution, “Penerapan Algoritma K-Means Dalam Mengclustering Kualitas Bibit,” vol. 1, no. 1, pp. 7–15, 2022.

[7] I. M. Pulungan, M. Fauzan, and A. P. Windarto, “Implementasi Algoritma K-Means Clustering dalam Menentukan Blok Tanaman Sawit Paling Produktif,” *Prosiding Seminar Nasional Riset Information* (Science), vol.1 no.3, pp. 338–348, September 2019, doi: 10.30645/senaris.v1i0.39.

[8] S. Hajar, A. A. Novany, A. P. Windarto, A. Wanto, and E. Irawan, “Penerapan K-Means Clustering Pada Ekspor Minyak Kelapa Sawit Menurut Negara Tujuan,” *Seminar Nasional Teknologi Komputer & Sains (SAINTEKS)*, pp. 314–318, 2020.

[9] A. N. Am and S. Defit, “Penentuan Mutu Kelapa Sawit Menggunakan Metode K-Means Clustering,” vol. 5, no. 3, pp. 1–9, 2019.

[10] T. Informatika, “Analisis Clustering untuk Mengetahui Tingkat Potensi Tanaman Kelapa Sawit Berdasarkan Luas Tanaman Menggunakan Algoritma K-Means pada Badan Pusat,” vol. 3, no. 1, pp. 110–123, 2019.

[11] Y. I. Febiola, I. Cholissodin, and C. Dewi, “Peramalan Hasil Panen Kelapa Sawit Menggunakan Metode Multifactors High Order Fuzzy Time Series yang Dioptimasi dengan K-Means Clustering (Studi Kasus : PT . Sandabi Indah Lestari Kota Bengkulu),” vol. 3, no. 12, 2019.

[12] W. E. Sari, M. Muslimin, A. Franz, and P. Sugiartawan, “Deteksi Tingkat Kematangan Tandan Buah Segar Kelapa Sawit dengan Algoritme K-Means,” *SINTECH (Science Inf. Technol. J.*, vol. 5, no. 2, pp. 154–164, 2022, doi: 10.31598/sintechjournal.v5i2.1146.

[13] and E. M. Pratama, Faiq Husain, Agung Triayudi, “Data mining k-medoids dan k-means untuk pengelompokan potensi produksi kelapa sawit di indonesia,” vol. 07, pp. 1294–1310, 2022.

[14] R. B. Ardi, F. E. Nastiti, and S. Sumarlinda, “Algoritma K-Means Clustering untuk Segmentasi Pelanggan (Studi Kasus : Fashion Viral Solo),” vol. 9, no. 1, pp. 124–131, 2023.

[15] N. Mirantika, “Penerapan Algoritma K-Means Clustering untuk Pengelompokan Penyebaran Covid-19," *Jurnal Nuansa Informatika,* Volume 15 Nomor 2 , Juli 2021, vol. 15, pp. 92–98, 2021.

[16] E. F. Himmah, M. Widyaningsih, and M. Maysaroh, “Identifikasi Kematangan Buah Kelapa Sawit Berdasarkan Warna RGB Dan HSV Menggunakan Metode K-Means Clustering,” *J. Sains dan Inform.*, vol. 6, no. 2, pp. 193–202, 2020, doi: 10.34128/jsi.v6i2.242.