ISSN: 2581-1274


# Clustering of Palm Oil Production Results with C-Means Algorithm in West Aceh Regency

Sanusi[a,1,*], Ilham Juliwardi[b,2], Muhammad Ardiansyah[c,3], Nica Astrianda[d,4], Ana Elviajakfar[e,5]

[a,b,c,d,e] *Universitas Teuku Umar, Meulaboh-Tapaktuan Street, Aceh Barat, 23615, Indonesia*
[1] *sanusi@utu.ac.id*;* [2] *Ilhamjuliwardi@utu.ac.id;* [3] *m.ardiansyah@utu.ac.id*
[4] *nicaastrianda@utu.ac.id;* [5] *anaelviajakfar@utu.ac.id*

ARTICLE INFO

*Article history:*
Accepted

*Keywords:*
Land Area
Production Result
C-Means Clustering

ABSTRACT

The area and production of oil palm in West Aceh continues to experience a significant increase, because along with the expansion of oil palm plantations in the local area, it is very influential on employment and improving the economy of the community. This study aims to group the results of oil palm production and planted area by applying the data mining technique method with the C-Means Clustering algorithm. The data used in this study in the form of area and palm oil production results were obtained from the West Aceh catalog (BAPPEDA) and the Central Statistics Agency from 2005 to 2020 in 12 sub-districts of West Aceh Regency. The results of the study showed that the high cluster (cluster 1) contained 1 sub-district, namely Kaway XVI, the medium cluster (cluster 0) there were 7 sub-districts namely Bubon, Johan Pahlawan, Panten Reu, Samatiga, Sungai Mas, Woyla Barat, and Woyla Barat, while the cluster low (cluster 2), there are 4 sub-districts, namely Arongan Lambalek, Meureubo, Pante Ceureumen, and Woyla.



## I. Introduction

Palm oil is one of the most profitable trade crops and has become one of the highest expansion values in Indonesia compared to other crop commodities. In 2016, the area of oil palm plantations in Indonesia was 11,201,465 hectares and in 2021 it reached 15,081,021 hectares, during the last 5 years the area of oil palm plantations in Indonesia increased by 34.5%. The area of commodity plantations is one of the largest contributors to Indonesia's foreign exchange, which can be seen from the value of exports. The total value of plantation exports reached US$ 25.38 billion or equivalent to Rp. 359.14 trillion. The plantation area is predicted to continue to increase in line with industrial developments and the fulfillment of vegetable oils in the world [1]. The area of oil palm plantations in Indonesia continues to grow rapidly, as does the production and export of palm oil. Its production in the form of CPO (crude palm oil) and CPKO (crude palm kernel oil) increased 17 times from 0.85 million tons to 14.4 million tons. Indonesia is currently the second largest producer of palm oil (CPO) and is predicted to be number one in 2010 (even the government is targeting 2008) in the world, surpassing Malaysia [2]. Along with the expansion of oil palm plantations, the production of palm oil in Indonesia also continues to increase. The results of palm oil production in 2016 produced 35000 tons (MT) to reach 44500 tons (MT) in 2021. Indonesian [3]. The area of land for palm oil production in Aceh is 129,169 hectares with a total production of 1,070,157 tons in 2019. Most of Aceh Province is a producer of palm oil spread over several districts. West Aceh Regency with a land area of 4,978 hectares, while the total production of palm oil from plantations is 75,435 tons, the results from community plantations are 13,518 tons [4]. Products produced from oil palm plantations in the form of palm oil or often called CPO and palm kernel oil or often called PKO. These products can be

DOI: 10.31572/inotera.Vol7.Iss1.2022.ID168 W : http://inotera.poltas.ac.id | E : inotera@poltas.ac.id

reprocessed for the purposes of several food and non-food industries, even as a source of renewable energy [5].

Based on field studies, it can be seen that the area and yield of oil palm production in the West Aceh continues to experience a significant increase, because along with the expansion of oil palm plantations in the local area, it is very influential on employment and can improve the community's economy and can encourage strengthening the regional economic structure, especially the West Aceh region. In addition to having a positive impact, plantation expansion also has a negative impact, namely the expansion and exploitation of plantations without legality which ultimately destroys ecosystem diversity and causes deforestation [6].

Clustering is a data mining technique used to partition a set of data objects into subsets. So that objects in a cluster are similar to each other, but different from objects in other clusters. The set of clusters resulting from cluster analysis is referred to as clustering. Different clustering methods can result in different groupings on the same data set. Object partitioning is not done by humans, but by a clustering algorithm [7].Therefore, it is necessary to do clusters or grouping of oil palm production, it is useful to know the most productive and unproductive oil palm plant blocks and can group productive oil palm plant types so that it can increase production and community welfare [8-10].

Previous studies such as cluster analysis of oil palm replanting locations. The aim of this research is to group potential oil palm producing areas to determine whether large or small areas of palm oil production are produced. The results showed that the C-Means method can be used for grouping priority areas for oil palm replanting in Bengkulu province [11]. Furthermore, the discovery of non-optimal production results of oil palm data processing using the C-Means algorithm. This study aims to determine which provinces are not yet optimal in producing palm oil. The results of the mapping can be seen that the Province is not optimal at 67%, optimal at 25% and more than optimal at 8% [12]. Furthermore, the use of the C-Means algorithm in mapping the potential yield of PTPN IV Marihat's palm oil production. This study tries to group potential fruit-producing locations to find out which locations produce large or small amounts of oil palm fruit. The results showed that C1 (highest) was 14 Harvest Block data, and C2 (lowest) was 11 Harvest Block data [13]. From several studies, it has been mentioned that the k-mean algorithm can be used in data grouping, because data from clusters with the same characteristics will be grouped in the same cluster, while different data will be grouped in other clusters.

This study aims to group the results of oil palm production and planted area by applying the data mining technique method with the C-Means Clustering algorithm. With this grouping, it is hoped that related parties and the public can find out more information and can be used to set strategies in increasing oil palm production and area in the South West Coast-West Aceh Regency.

## II. Method

The application of data mining techniques with the C-Means clustering algorithm in this study consisted of several stages to obtain a grouping of areas and yields of high and low potential palm oil production. Based on Figure 1, it can be described the steps carried out in this study. The data used to classify the results of oil palm production is secondary data consisting of land area data and production yields obtained from the Aceh Barat catalog in Figures in Collaboration with the Regional Development Planning Agency (BAPPEDA) with the Central Statistics Agency for Aceh Barat [14], via the website http: //acehbaratkab.bps.go.id Data on land area and palm oil production used in this study for 16 years starting from 2005 to 2020. After the data is collected, the next step is to carry out a pre-process starting from checking the completeness of the data until the data to be clustered is no longer added and subtraction, then assign a label to each variable. The segmentation of land area and palm oil production is carried out using Rapid Miner Studio V7.1 software. The stages of this research are as follows:

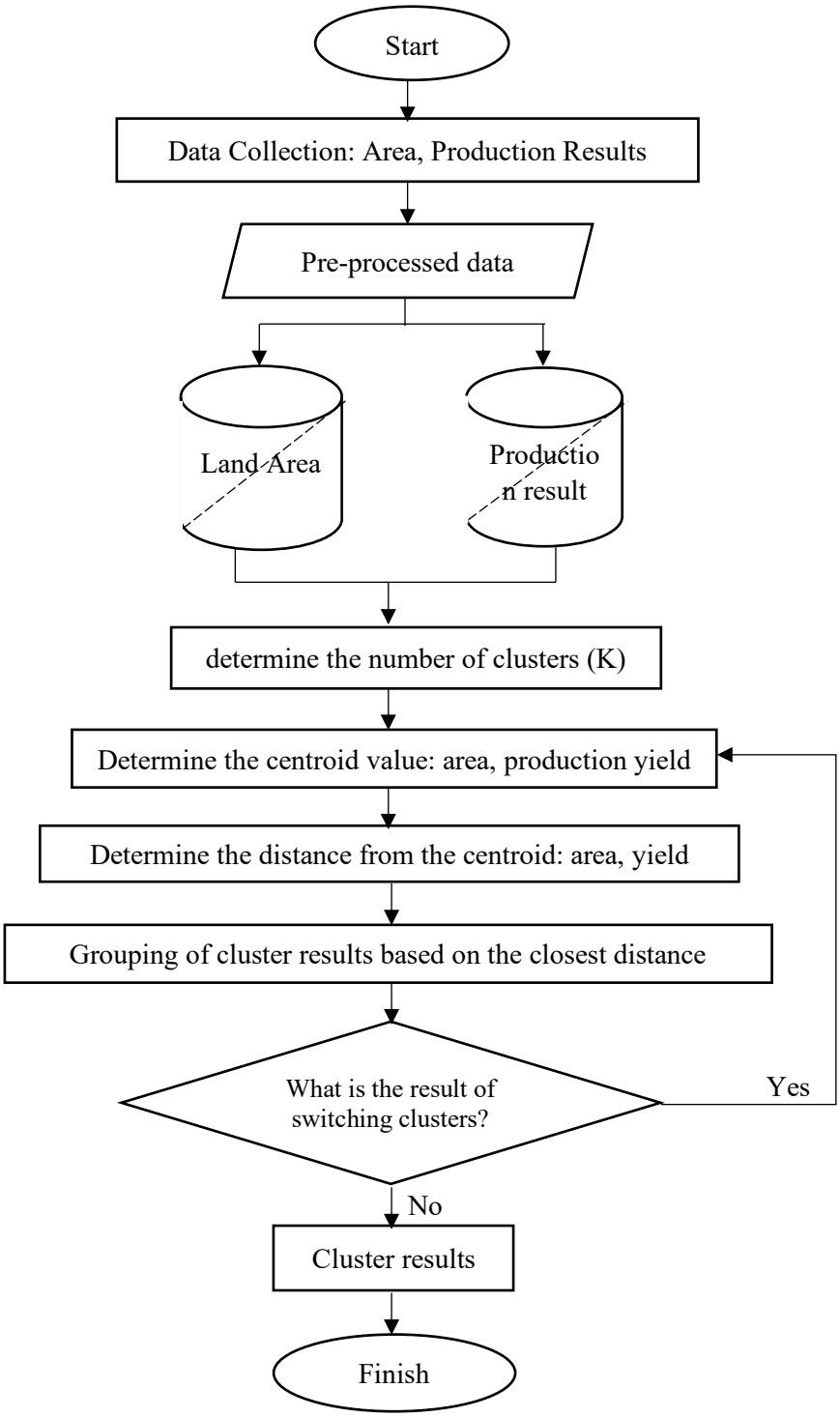


Figure 1. Research Flowchart

The data mining technique with the C-Means Clustering algorithm will be carried out after the data is transformed and attributes are given to each column, namely sub-district attributes, area, and production results, so each year it has 12 sub-districts for 16 years.

The following are the steps taken by the author in classifying the results of oil palm production using the C-Means algorithm [13][15-16]:

1. Determine the number of clusters as many as K clusters, this cluster is done by taking the largest value for high clusters (C1), medium clusters for (C2), and low clusters for (C3).

2. Determine the initial centroid (cluster center) value which is determined randomly based on the value of the data variable in the cluster as much as the predetermined k value.

3. Calculate the distance of each data to the center of the cluster using the Euclidean Distance formula:

$$dist = \sqrt{\sum_{k=1}^{n} (p_k - q_k)^2}$$

Description:
Dist is the distance between a point and the center point, n is the number of dimensions (attributes), pk and qk are k attributes (components) or data objects p and q, respectively.

1. Determine the cluster position of each data on the results of oil palm production based on the closest distance to the cluster center.
2. Calculate the iteration result centroid value by checking the condition of stopping the iteration process, if the iteration result centroid value with the previous centroid value is the same or the centroid value is optimal then the iteration process stops. However, if the centroid value of the iteration results is not the same or not optimal, then the iteration process continues in the next iteration.

## III. Results and Discussion

The data used in this study consists of the area and results of oil palm production from 2005 to 2020 in 12 sub-districts of West Aceh Regency. The results of the selection of data on area and palm oil production results are as shown in table 1 and 2 below. The table, show results of the data selection can be shown. The data generated at the data selection stage will be transformed to produce the right data when used in the data mining process with the C-Means clustering algorithm. Data transformation is carried out by selecting attributes as shown in table 3.

Table 1. Wide Area

| No | Sub-District | 2005 | 2006 | 2007 | 2008 | … | 2020 |
|---|---|---|---|---|---|---|---|
| | | Area | Area | Area | Area | … | Area |
| 1 | Arongan Lambalek | 106 | 106 | 376 | 376 | … | 1353 |
| 2 | Bubon | 151 | 160 | 225 | 225 | … | 581 |
| 3 | Johan Pahlawan | 25 | 25 | 25 | 25 | … | 297 |
| 4 | Kaway XVI | 1712 | 1765 | 2084 | 2084 | … | 2984 |
| 5 | Meureubo | 751 | 754 | 754 | 754 | … | 1060 |
| 6 | Pante Ceureumen | 324 | 324 | 422 | 422 | … | 664 |
| 7 | Panton Reu | 51 | 55 | 60 | 60 | … | 408 |
| 8 | Samatiga | 40 | 53 | 53 | 53 | … | 264 |
| 9 | Sungai Mas | 146 | 146 | 146 | 146 | … | 503 |
| 10 | Woyla | 294 | 294 | 476 | 476 | … | 1053 |
| 11 | Woyla Barat | 181 | 181 | 181 | 181 | … | 629 |
| 12 | Woyla Timur | 162 | 162 | 229 | 229 | … | 724 |

Table 2. Production Results

| No | Sub-District | 2005 | 2006 | 2007 | 2008 | … | 2020 |
|---|---|---|---|---|---|---|---|
| | | Output | Output | Output | Output | … | Output |
| 1 | Arongan Lambalek | 237.31 | 237.31 | 848.75 | 1367.80 | … | 1734.18 |
| 2 | Bubon | 703.87 | 703.87 | 1389.75 | 1365.40 | … | 642.74 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 3 | Johan Pahlawan | 103.58 | 103.58 | 242.75 | 251.00 | … | 415.80 |
| 4 | Kaway XVI | 8042.27 | 8042.27 | 16244.50 | 18521.80 | … | 4177.60 |
| 5 | Meureubo | 1257.96 | 1257.96 | 6022.50 | 6646.80 | … | 1484.00 |
| 6 | Pante Ceureumen | 1014.16 | 1014.16 | 2476.00 | 3116.80 | … | 873.60 |
| 7 | Panton Reu | 231.42 | 251.50 | 351.52 | 390.80 | … | 452.20 |
| 8 | Samatiga | 80.20 | 80.20 | 386.25 | 448.20 | … | 369.60 |
| 9 | Sungai Mas | 395.48 | 395.48 | 1694.25 | 1745.40 | … | 574.40 |
| 10 | Woyla | 985.30 | 985.30 | 3406.25 | 4187.60 | … | 1453.90 |
| 11 | Woyla Barat | 398.90 | 398.90 | 1200.25 | 1738.80 | … | 644.00 |
| 12 | Woyla Timur | 299.52 | 299.52 | 970.75 | 1860.80 | … | 645.40 |

Table 3. Results of Data Transformation

| Sub-district | Year | Area (Ha) | Output Production (Kg) |
|---|---|---|---|
| Arongan Lambalek | 2005 | 106.00 | 237.31 |
| Bubon | 2005 | 151.00 | 703.87 |
| Johan Pahlawan | 2005 | 25.00 | 103.58 |
| Kaway XVI | 2005 | 1712.00 | 8042.27 |
| Meureubo | 2005 | 751.00 | 1257.96 |
| Pante Ceureumen | 2005 | 324.00 | 1014.16 |
| Panton Reu | 2005 | 51.00 | 231.42 |
| Samatiga | 2005 | 40.00 | 80.20 |
| Sungai Mas | 2005 | 146.00 | 395.48 |
| Woyla | 2005 | 294.00 | 985.30 |
| Woyla Barat | 2005 | 181.00 | 398.90 |
| Woyla Timur | 2005 | 162.00 | 299.52 |
| Arongan Lambalek | 2006 | 106.00 | 237.31 |
| Bubon | 2006 | 160.00 | 703.87 |
| … | … | … | … |
| Woyla Timur | 2020 | 724.00 | 645.40 |

Table 3 shows the results of data transformation based on the order of sub-district, year, area (ha), and production yield (kg). After that, it is necessary to find the average area (ha) and production yield (kg) in each sub-district based on the order of the year before segmentation using the C-Means clustering algorithm with rapid miner studio. The results of the average search for each attribute to be segmented are as shown in table 4 below:

Table 4. Average Area and Production Results

| Sub-district | Area | Output Production |
|---|---|---|
| Arongan Lambalek | 776.67 | 2813.79 |
| Bubon | 331.54 | 1854.01 |
| Johan Pahlawan | 150.69 | 362.46 |
| Kaway XVI | 2413.63 | 16377.93 |
| Meureubo | 906.44 | 5055.88 |
| Pante Ceureumen | 509.14 | 2954.28 |
| Panton Reu | 152.2 | 484.41 |

| | | |
|---|---|---|
| Samatiga | 144 | 845.53 |
| Sungai Mas | 253.69 | 1425.26 |
| Woyla | 679 | 3860.95 |
| Woyla Barat | 307 | 1392.51 |
| Woyla Timur | 340.31 | 1711.01 |

Table 4 is a dataset of 12 sub-districts and each sub-district has an averaged area and production results which will later be processed using the C-Means Clustering algorithm. Segmentation is done using Rapid Miner software.

## Design Rapidminer

The first thing to do is make a design by importing the data contained in Table 2 by choosing the C-Means and performance algorithm. When importing data, several new operators must be selected, as shown in Figure 2 below:

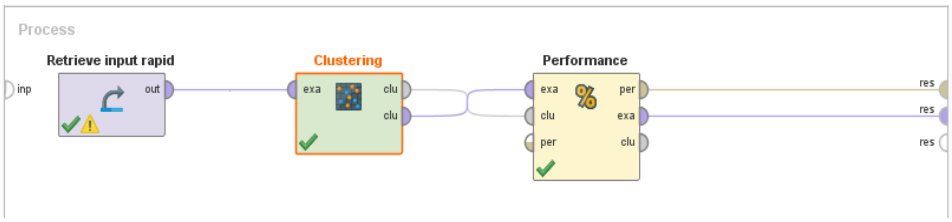


Figure 2. Import Data

Based on Figure 2, it can be seen the results of the C-Means clustering algorithm design process starting from importing data, choosing the C-Means operator, and modeling performance when clustering palm production. Segmentation is determined as much as C = 3 which can provide information on the high group (C1), medium group (C2), and low group (C3). To get the initial centroid value or the initial center point of the cluster, randomly selected and automatically determined in the Rapidminer software, while measuring the centroid distance to the initial data using Euclidean Distance with a number of iterations of 10 until the cluster position does not change. The attributes used as labels are sub-district data, area, and production results.

Table 5. Data Centroid

| Variabel | Cluster 0 | Cluster 1 | Cluster 2 |
|---|---|---|---|
| Area | 239.91857142857143 | 2413.63 | 717.8125 |
| Production | 1153.5985714285714 | 16377.93 | 3671.2250000000004 |

Based on table 5, it can be seen that the centroid value obtained in the 10 iteration is determined automatically in rapidminer, namely the centroid value in each cluster based on the area and production results used according to the table centroid.

Table 6. Statistical value of each variable

| Variable | Min | Max | Averafe | Standard Deviation |
|---|---|---|---|---|
| Area | 144 | 2413.630 | 580.359 | 631.007 |
| Production | 362.460 | 16377.930 | 3261.502 | 4360.132 |

Based on table 6, it can be seen that the area of oil palm planting for 16 years in 12 sub-districts of West Aceh Regency obtained a minimum value of 144 (ha), a maximum of 2413,630 (ha), an average of 580,359 (ha), and a standard deviation of 631,007 (ha). Meanwhile, the minimum production yield is 362,460 kg, the maximum is 16377,930 (kg), the average is 3261,502 (kg), and the standard deviation is 4360,132 (kg). Based on the modeling results obtained, it can be concluded that the wider the planted area, the higher the production yield.

Table 7. Results of Cluster Model

| Cluster | Sub-district |
|---|---|
| Cluster 0 | Bubon, Johan pahlawan, Panten reu, Samatiga, Sungai mas, Woyla barat, Woyla timur |
| Cluster 1 | Kaway XVI |
| Cluster 2 | Arongan lambalek, Meureubo, Pante ceureumen, Woyla |

Based on table 7, it can be seen that the results of the C-Means clustering process obtained three cluster models, from these results it can provide information that the high cluster (cluster 1) has 1 sub-district, namely Kaway XVI, the medium cluster (cluster 0) is 7 sub-districts namely Bubon, Johan Pahlawan. , Panten reu, Samatiga, Sungai mas, Woyla west, and Woyla east, while the low cluster (cluster 2) contains 4 sub-districts, namely Arongan lambalek, Meureubo, Pante ceureumen, and Woyla. While the graph of the results of the cluster as shown in Figure 3 below:

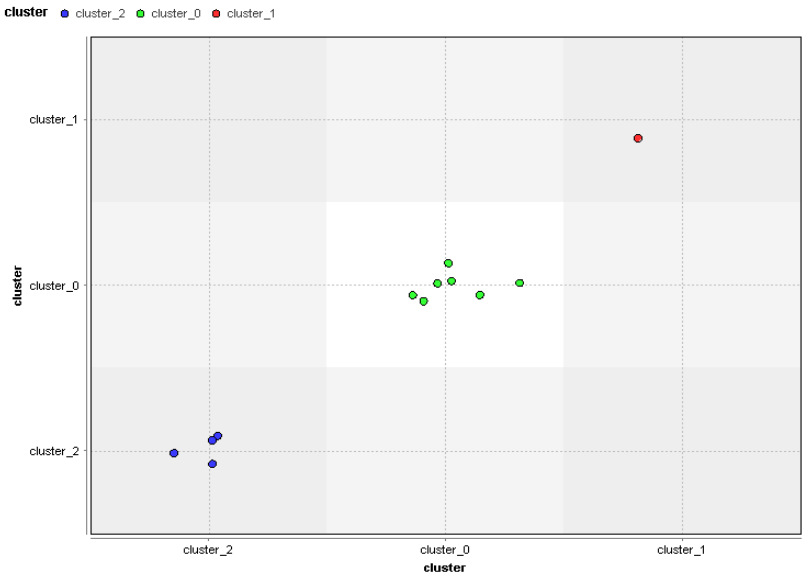


Figure 3. Cluster graph

Based on Figure 3, the computational results of the k-mean clustering process show the number of items in each cluster of palm oil production in 12 sub-districts in the West Aceh Regency. Representation of C-Means Clustering, it can be seen that cluster 1 is shown in red which is a group of oil palm plantations that have a large area and oil palm production is included in the high category, namely in Kaway XVI sub-district. The results of the field review are that the majority of the people earn from oil palm plantations and have large oil palm plantations. In addition, in this sub-district there is also a palm oil company PT. Karya Tanah Subur (KTS). Furthermore, cluster 0 is shown in green which is a group of oil palm plantations that have an area that is included in the medium category and the results of oil palm production are also included in the medium category, namely in the districts of Bubon, Johan Pahlawan, Panten Reu, Samatiga, Sungai Mas, Woyla Barat, and Woyla Timur. While cluster 2 is the lowest cluster result shown in blue in Arongan lambalek, Meureubo, Pante ceureumen, Woyla sub-districts where the sub-district has a small land area and low production yields.

## IV. Conclusion

Based on the results of data mining modeling with the C-Means clustering algorithm, it can be concluded that the C-Means algorithm can be implemented on mapping data on the area and production of palm oil in the West Aceh region. The results of this study found that the high cluster (cluster 1) contained 1 sub-district namely Kaway XVI, the medium cluster (cluster 0) was 7 sub-districts namely Bubon, Johan Pahlawan, Panten Reu, Samatiga, Sungai mas, Woyla Barat, and

Woyla Timur, while In the low cluster (cluster 2), there are 4 sub-districts, namely Arongan lambalek, Meureubo, Pante ceureumen, and Woyla.

## References

[1] Direktorat Jenderal Perkebunan, 2021. Statistik Perkebunan Unggulan Nasional 2019-2021. www.ditjenbun.pertanian.go.id.

[2] Samhadi, S.H. 2006. Ironi Sawit dan Ambisi Nomor Satu Dunia. Komisi Pengawas Pesaingan Usaha Republik Indonesia.

[3] Palm Oil Production by Country in 1000 MT," Index Mundi, accessed March 02 2022, https://www.indexmundi.com/agriculture/?commodity=palm-oil&country=id&graph=production.

[4] Asnawi, Yusra, Aiyub, Amru. 2013. Pengembangan Coloring Economic Model suatu Strategi Kemiteraan antara Sektor Karet dan Kelapa Sawit sebagai Penggerak Ekonomi dalam Upaya Mengurangi Kemiskinan Di Provinsi Aceh. Jurnal Ekonomi, Manajemen dan Akuntansi Vol 1 No. 1. ISSN 2013 23030542.

[5] Verneau, F. Barbera, L. Amato, M. Sodano, V. (2019). Consumers' concern towards palm oil consumption an empirical study on attitudes and intention in Italy. Br. Food J., vol. 121, no. 9, pp. 1982–1997, 2019.

[6] Hutabarat, S. (2019). Optimalisasi Pemanfaatan Lahan Perkebunan Kelapa Sawit di Riau. Unri Conf. Ser. Agric. Food Secur., vol. 1, pp. 46–57, 2019.

[7] Han, J. Kamber, M. Pei, J. (2012). Data Mining: Concepts and Techniques. University of Illinois at Urbana–Champaign, Publishers of Elsevier. ISBN 978-0-12-381479-1.

[8] Pulungan, M. Fauzan, M. Windarto, A.P. (2009). Implementasi Algoritma C-Means Klustering dalam Menentukan Blok Tanaman Sawit Paling Produktif. in Prosiding Seminar Nasional Riset Information Science (SENARIS), 2019, no. September, pp. 338-348.

[9] Saragih, R,A,D. Safii, M. Tambunan, H.S. (2019). Penerapan Metode C-Means Klustering Untuk Mengelompokkan Kelapa Sawit Produktif. in Prosiding SiManTap: Seminar Nasional Matematika dan Terapan, pp. 362–370.

[10] Pasaribu, D. Damanik, F.I.S. Irawan, E. Tambunan, H.S. Kunci, K. 2021. Memanfaatkan Algoritma C-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.

[11] Fauzi, E. Harta, L. Oktavia, R. Sudarmansyah, Astuti, HB. Firison, J. Ishak, A. Yuliasari, S. Sastro, Y. 2020. Kluster Analysis of Determining the Location of Oil Palm Replanting in Mukomuko District. Agritropica: Journal of Agricultural Science. 3 (1): 38-45. Doi:https://doi.org/10.31186/Jagritropica.3.1.38-45.

[12] Shyntya, Y. Marleen O. 2019. Finding of Non-Optimal Production Results for Palm Oil Data Processing Using C-Means Algorithm. Journal of Software Engineering & Intelligent Systems; Volume 4, Issue 2, ISSN: 2518-8739.

[13] Pasaribu, D.F. Damanik, I.S. Irawan, E. Suhada. Tambunan, H.S. 2021. Memanfaatkan Algoritma C-Means dalam Memetakan Potensi Hasil Produksi Kelapa Sawit PTPN IV Marihat. BIOS: Jurnal Teknologi Informasi dan Rekayasa Komputer Vol. 2, No. 1, hlm. 11-20, ISSN: 2722-0850.

[14] Aceh Barat dalam Angka: Badan Perencanaan Pembangunan Daerah (BAPPEDA) dengan Badan Pusat Statistik Aceh Barat: website http://acehbaratkab.bps.go.id/ Diakses: 20/03/2022.

[15] Unnati, R. Raval. Jani, C. 2016. Implementing & Improvisation of C-Means Klustering Algorithm. International Journal of Computer Science and Mobile Computing, Vol.5 Issue.5, pg. 191-203.

[16] Sanusi, Husna, J. 2020. Utilization of Rapidminer using the C-Means Klustering Algorithm for Classification of Dengue Hemorrhagic Fever (DHF) Spread in Banda Aceh City. Jurnal Inovasi Teknologi dan Rekayasa. ISSN: 2581-1274 Vol. 5, No. 2, July-December 2020, pp. 146-15