A Oil Palm Harvest Grouping Using K-Medoids Algorithm

Dessy Dwi Angraini¹, M. Safii¹, Fitri Anggraini²
¹Informatics Engineering Study Program, STIKOM Tunas Bangsa, Pematangsiantar
²Computerized Accounting Study Program, Amik Tunas Bangsa, Indonesia

Article Info

Article history:
Received Jun 20, 2021
Revised Jul 22, 2021
Accepted Sep 17, 2021

Keywords:
Oil Palm
Janjang
Data Mining
K-Medoids
RapidMiner

ABSTRACT

Oil palm (Elaies Guinensis Jacq) is one of the important industrial crops producing cooking oil, industrial oil, and fuel. Indonesia is the largest palm oil producer in the world. The rest of the processing of oil palm fruit is called janjang. Janjang also serves to be used as compost. The data that is processed in this research is the harvest data at PT. Surya Intisiraiyara Mandau. Data mining is the process of looking for patterns or information in selected data using certain techniques or methods. The processing steps are grouped using the K-Medoids method and then the data will be processed using RapidMiner tools. Where this grouping is done to minimize the amount of similarity of data and appropriate so that it becomes more valid data. This study aims to simplify the grouping of harvest data based on high, medium and low clusters.

1. INTRODUCTION

Oil palm (Elaies Guinensis Jacq) is one of the important industrial crops producing cooking oil, industrial oil, and fuel [1]. Indonesia is the world's largest producer of palm oil [2][3][4]. The most important part to be processed from the oil palm plant is the fruit. The rest of the processing of oil palm fruit (janjang) is very potential which can be used as compost [5][6][7]. As one of the largest agricultural export commodities in Indonesia, palm oil has an important role as a source of foreign exchange and large tax earners.

Several previous studies that were used as references in this study conducted by [8] explained that forecasting oil palm yields consisted of several factors, namely monthly oil palm yields, land area, plant age, and principal amount. Palm oil. Subsequent research by [9][10] The process of identifying the maturity of oil palm fruit using the K-Means Clustering method is able to recognize oil palm fruit image objects based on the level of maturity, namely raw, moderately ripe, and ripe and the results of the identification of oil palm fruit maturity with K-means Clustering algorithm obtained an accuracy rate of 79.16% for test data and a level 2 accuracy for training data of 50%, so that the total accuracy of both is 64.58%. Harvesting is one of the important activities in the management of mature oil palm plantations [11]. In addition to planting material (seeds) and plant
maintenance, harvesting is also an important factor in achieving productivity. In making harvest data at PT. Intisariraya Solar Saber, Riau where the processing of data not harvest bunch of data grouped by cluster of high, medium, and low on-AFD3 AFD1 per month. Therefore, the author uses the k-medoids method as a method for classifying harvest data. The purpose of this grouping is to facilitate PT. Surya Intisariraya Mandau, Riau in knowing the number of harvest bunch in AFD1-AFD3 which consists of several blocks that get results bunch production to cluster high, cluster medium, and the cluster is low.

The object of this research uses data from the 2020 harvest at PT. Surya Intisariraya Mandau, Riau. The concept that will be used in this research is to use the concept of data mining. Data mining technique is a program for processing large volumes of data and has a high speed and can be used as a source of information and used to draw conclusions related to a research to be processed \[12\][13][14]. In the process, data mining uses statistical, mathematical, artificial intelligence and machine learning techniques that function to extract and identify useful information and related knowledge from various large databases.

## 2. RESEARCH METHOD

The design of this study was first carried out by observing data and analyzing existing problems. After that the data will be processed through the calculation process and follow the steps of the k-medoids algorithm calculation process. Furthermore, the results of the grouping will be applied to Rapidminer to see the results. The results of grouping the data on the harvest of janjang are the development of science which will later be able to provide solutions to problem solving in PT. Surya Intisariraya Mandau, Riau. Flowchart is a technique used to explain aspects of information systems clearly, precisely and logically.

![Flowchart K-medoids](image)

The following are some examples of long harvest data samples per month in 2020, where the total number of perennial harvest data is 140 data.

### Table 1. Data Sample
(Source: PT. Surya Intisariraya Mandau)

<table>
<thead>
<tr>
<th>AF</th>
<th>Blok</th>
<th>Large</th>
<th>Harvest Janjang (kg) / Month</th>
</tr>
</thead>
</table>

---

_A Oil Palm Harvest Grouping Using K-Medoids Algorithm (Dessy Dwi Angraini, et al)_
At this stage, the data transformation/ preprocessing process is carried out, where data showing numbers < 1000 = 1, > 1000 = 2, > 2000 = 3, > 3000 = 4, > 4000 = 5. The collected data will be processed into 3 clusters. with several stages and the calculation process using the rapidminer version 5.3 application.

Table 2. Sample Data Used

<table>
<thead>
<tr>
<th>AFD</th>
<th>Blok</th>
<th>Harvest Janjang (kg) / Month</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Jan</td>
</tr>
</tbody>
</table>
| 1   | S34  | 2.09| 2.19| 1.761| 3.456.| 3.22| 3.6 | 3.246| 3.331| 3.2738.
| 1   | S30  | 2.43| 1.86| 1.635| 3.028.| 2.71| 3.6 | 3.432| 3.716| 2.783.
| 1   | S31  | 2.277| 1.80| 2.105| 2.85. | 3.2 | 3.71| 3.778| 3.120| 2.617.
| 1   | U36  | 2.723.0| 1.864| 9.0 | 2.393. | 3.7 | 3.4 | 4.056.| 3.856.| 3.244. 2.723.
| 2   | P32  | 2.36| 9.0 | 2.189.| 2.971.| 3.035| 3.281| 3.05 | 3.420| 3.409. |

Harvest Janjang (kg) / Month

<table>
<thead>
<tr>
<th>AFD</th>
<th>Blok</th>
<th>Harvest Janjang (kg) / Month</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Jan</td>
</tr>
<tr>
<td>2</td>
<td>P33</td>
<td>1.743</td>
</tr>
<tr>
<td>2</td>
<td>P34</td>
<td>1.783</td>
</tr>
<tr>
<td>3</td>
<td>B07</td>
<td>2.337.</td>
</tr>
<tr>
<td>3</td>
<td>B08</td>
<td>2.285.</td>
</tr>
</tbody>
</table>

RESULTS AND DISCUSSIONS

To solve the problem in this research, data mining techniques with the k-medoids algorithm are used as follows: [15],[16],[17],[18],[19],[20]:

\[ D_{ij} = \sqrt{\sum_{a=1}^{p} (x_{ia} - x_{ja})^2} = \sqrt{(x_i - x_j)'(x_i - x_j)} \]

3.1 Solution with K-Medoids

The K-Medoids method used in the clustering process of harvesting janjang data obtained from PT. Surya Intisariraya Mandau. The steps in completing manual data mining calculations using the k-medoids clustering method are as follows:

a. **Initialization of k-medoids cluster center (number of clusters)**

The number of clusters used in this study consisted of C2/high, C1/medium, C0/low. The k-medoids algorithm begins with the initial determination of the cluster center by randomly selecting among the objects in the panen janjang data.

**Table 3. Initial Cluster Center Initialization**

<table>
<thead>
<tr>
<th>Blok N31 (C1)</th>
<th>Blok S34(C2)</th>
<th>Blok N30 (C3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jan</td>
<td>Feb</td>
<td>Mar</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
</tbody>
</table>

b. Calculate the value of the closest distance with the Euclidean Distance equation. To calculate the distance between the centroid point with the point of each object using Euclidean Distance. Then the manual calculation to calculate the distance of each object with the initial medoids is as follows:

\[ D_{U31, C1} = \sqrt{(1 - 1)^2 + (1 - 2)^2 + (1 - 1)^2 + (1 - 1)^2} = 3 \]

\[ D_{S30, C2} = \sqrt{(2 - 2)^2 + (2 - 2)^2 + (2 - 2)^2 + (3 - 3)^2} = 2 \]
\[ D_{Nab, C_3} = \sqrt{(1 - 2)^2 + (2 - 2)^2 + (1 - 1)^2 + (2 - 2)^2 + (2 - 2)^2} \]
\[ \quad = \sqrt{2} \]

\begin{table}[h]
\centering
\begin{tabular}{|c|c|c|c|c|}
\hline
C1 & C2 & C3 & Nearest Distance & Cluster  \\
\hline
4  & 0  & 3  & 0  & 2  \\
5  & 2  & 3  & 2  & 2  \\
5  & 3  & 4  & 2  & 2  \\
3  & 2  & 3  & 2  & 2  \\
6  & 3  & 4  & 3  & 2  \\
3  & 7  & 4  & 3  & 1  \\
3  & 2  & 3  & 2  & 2  \\
6  & 3  & 5  & 3  & 2  \\
4  & 2  & 3  & 2  & 2  \\
3  & 6  & 4  & 3  & 1  \\
\hline
\end{tabular}
\caption{Medoids Distance Iteration 1}
\end{table}

c. Calculating Cost Value, the cost value is obtained from the total sum of the medoid closest distance values.
d. Repeat the previous steps with the new medoids center value. The \textit{k-medoids} algorithm begins with the initial determination of the cluster center by randomly selecting among the objects in the long harvest data for each object as a candidate for new medoids.

\begin{table}[h]
\centering
\begin{tabular}{|c|c|c|c|c|c|c|c|c|c|c|}
\hline
\hline
Blok V33 (C1) & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 2 & 1 & 1 & 1 & 1  \\
Blok U35 (C2) & 2 & 2 & 3 & 3 & 2 & 3 & 4 & 3 & 3 & 2 & 3 & 2  \\
Blok N30 (C3) & 2 & 2 & 2 & 2 & 2 & 2 & 3 & 3 & 2 & 3 & 2 & 1  \\
\hline
\end{tabular}
\caption{New Medoids Cluster Center Iteration 2}
\end{table}

\[ D_{Bnk, C_1} = \sqrt{(1 - 2)^2 + (1 - 2)^2 + (1 - 1)^2 + (1 - 1)^2} \]
\[ \quad = \sqrt{2} \]

\[ D_{T3k, C_2} = \sqrt{(3 - 2)^2 + (2 - 2)^2 + (2 - 3)^2 + (4 - 3)^2} \]
\[ \quad = 4 \]

\[ D_{V3k, C_3} = \sqrt{(1 - 2)^2 + (1 - 2)^2 + (1 - 2)^2 + (2 - 2)^2} \]
\[ \quad = 3 \]

\begin{table}[h]
\centering
\begin{tabular}{|c|c|c|c|c|}
\hline
C1 & C2 & C3 & Nearest Distance & Cluster  \\
\hline
6  & 2  & 3  & 2  & 2  \\
7  & 2  & 3  & 2  & 2  \\
7  & 2  & 3  & 2  & 2  \\
7  & 3  & 4  & 3  & 2  \\
\hline
\end{tabular}
\caption{New Medoids Distance Iteration 2}
\end{table}
e. Calculate the total deviation (S) by reducing the cost value in iteration 2 (new) with iteration 1 (initial). If the value of S < 0 then the processing is continued by using the new medoids center value. If the value of S > 0 or the value of the cost of the new iteration is greater than the value of the cost of the old iteration, the process is stopped so that the value of S is obtained:

\[
S = \text{new cost value} - \text{old cost value}
\]

\[
= 182 - 177 = 5
\]

Because the New Cost Value > Old Cost Value, the iteration is stopped and the cluster result is in iteration 1.

3.2 Results of Data Processing With RapidMiner

RapidMiner is software for data processing using data mining principles and algorithms [21],[22],[23],[24],[25].

![Cluster Data Results](image)

**Figure 2. Cluster Data Results**

In figure 2, it can be seen that the results of data clusters are in AFD1-AFD3 and each block has been grouped based on cluster 0, cluster 1, and cluster 2.

![Cluster Model Results](image)

**Figure 3. Cluster Model Results**

In figure 3, it can be seen the results of clustering that have been processed, where cluster 0 consists of 12 data, cluster 1 consists of 62 data, and cluster 2 consists of 6 data.
Figure 4. Centroid Table Results on RapidMiner

In figure 3.3, the results of the centroid table from the calculation of \( k \)-medoids using rapidminer tools have the same results as the results of manual calculations.

3.3 Final Result of K-Medoids and RapidMiner Tools

<table>
<thead>
<tr>
<th>Attribute</th>
<th>cluster_0</th>
<th>cluster_1</th>
<th>cluster_2</th>
</tr>
</thead>
<tbody>
<tr>
<td>AFD</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>BLOK</td>
<td>79</td>
<td>77</td>
<td>78</td>
</tr>
<tr>
<td>JAN</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>FEB</td>
<td>1</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>MAR</td>
<td>1</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>APR</td>
<td>1</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>MEI</td>
<td>1</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>JUN</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>JUL</td>
<td>1</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>AGST</td>
<td>2</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>SEP</td>
<td>2</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>OKT</td>
<td>2</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>NOV</td>
<td>2</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>DES</td>
<td>2</td>
<td>3</td>
<td>2</td>
</tr>
</tbody>
</table>

Tabel 6. Final Result of K-Medoids and RapidMiner Tools

<table>
<thead>
<tr>
<th>AFD</th>
<th>BLOCK</th>
<th>K-Medoids Algorithm Calculation</th>
<th>Testing With RapidMiner App</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>S34</td>
<td>Cluster_2</td>
<td>Cluster_1</td>
</tr>
<tr>
<td>1</td>
<td>S30</td>
<td>Cluster_2</td>
<td>Cluster_1</td>
</tr>
<tr>
<td>1</td>
<td>S31</td>
<td>Cluster_2</td>
<td>Cluster_1</td>
</tr>
<tr>
<td>1</td>
<td>S32</td>
<td>Cluster_2</td>
<td>Cluster_1</td>
</tr>
<tr>
<td>1</td>
<td>S33</td>
<td>Cluster_2</td>
<td>Cluster_1</td>
</tr>
<tr>
<td>1</td>
<td>S35</td>
<td>Cluster_2</td>
<td>Cluster_1</td>
</tr>
<tr>
<td>1</td>
<td>S36</td>
<td>Cluster_2</td>
<td>Cluster_1</td>
</tr>
<tr>
<td>1</td>
<td>T32</td>
<td>Cluster_2</td>
<td>Cluster_1</td>
</tr>
<tr>
<td>1</td>
<td>T33</td>
<td>Cluster_2</td>
<td>Cluster_1</td>
</tr>
<tr>
<td>1</td>
<td>T34</td>
<td>Cluster_2</td>
<td>Cluster_1</td>
</tr>
<tr>
<td>1</td>
<td>T35</td>
<td>Cluster_2</td>
<td>Cluster_1</td>
</tr>
<tr>
<td>1</td>
<td>T31</td>
<td>Cluster_2</td>
<td>Cluster_1</td>
</tr>
<tr>
<td>2</td>
<td>P34</td>
<td>Cluster_2</td>
<td>Cluster_1</td>
</tr>
<tr>
<td>2</td>
<td>P35</td>
<td>Cluster_2</td>
<td>Cluster_1</td>
</tr>
<tr>
<td>2</td>
<td>P36</td>
<td>Cluster_2</td>
<td>Cluster_1</td>
</tr>
<tr>
<td>2</td>
<td>P37</td>
<td>Cluster_2</td>
<td>Cluster_1</td>
</tr>
<tr>
<td>2</td>
<td>P34</td>
<td>Cluster_2</td>
<td>Cluster_1</td>
</tr>
<tr>
<td>3</td>
<td>B07</td>
<td>Cluster_2</td>
<td>Cluster_1</td>
</tr>
<tr>
<td>3</td>
<td>B08</td>
<td>Cluster_2</td>
<td>Cluster_1</td>
</tr>
<tr>
<td>3</td>
<td>B09</td>
<td>Cluster_2</td>
<td>Cluster_1</td>
</tr>
<tr>
<td>3</td>
<td>B10</td>
<td>Cluster_2</td>
<td>Cluster_1</td>
</tr>
<tr>
<td>3</td>
<td>B11</td>
<td>Cluster_2</td>
<td>Cluster_2</td>
</tr>
</tbody>
</table>

4. CONCLUSION

Based on the results of the research that has been done, the following conclusions can be drawn, namely, the clustering of harvest data using the \( k \)-medoids clustering method resulted in 3 clusters, namely cluster 1 (low) consisting of 12 data, cluster 2 (medium) consisting of 62 data, and cluster 3 (high) consisting of 6 data. Knowing the process of testing the \( k \)-medoids method on harvested harvest data using rapidminer tools.
ACKNOWLEDGEMENTS

Thank you to the supervisors as lecturers at AMIK and STIKOM Tunas Bangsa Pematangsiantar who have assisted in the preparation of this research so that it has reached the publication stage. Thanks to all parties, especially PT. Intisariraya Mandau, Riau and it is hoped that this research can be used as a source of data that can assist in decision making. For the development and progress of this research title, we expect constructive input and criticism so that this research can be continued.

REFERENCES


A Oil Palm Harvest Grouping Using K-Medoids Algorithm (Desy Dwi Angraini, et al)


