Engineering and Technology Journal e-ISSN: 2456-3358

Volume 10 Issue 11 November-2025, Page No.-7923-7930

DOI: 10.47191/etj/v10i11.23, I.F. – 8.482

Clustering of Screw Press Machine Conditions using the Agglomerative Hierarchical Clustering

Irfan Dwi Prawira[1], Jasril[2], Suwanto Sanjaya[3], Lestari Handayani[4], Fitri Insani[5]

[1,2,3,4,5]Informatics Engineering, Faculty of Science and Technology, Sultan Syarif Kasim Riau State Islamic University.

ABSTRACT: The screw press machine is an important component in the palm oil processing process that functions to extract oil from palm kernels. Continuous use of the machine can reduce performance and disrupt the production process. Therefore, machine condition analysis is needed to support preventive maintenance strategies. One method that can be used is the clustering technique. Clustering is a technique for grouping data based on specific parameters to form classes with similar characteristics. This study applied the Agglomerative Hierarchical Clustering (AHC) method with a single linkage approach to group the conditions of screw press machines based on data obtained from PT. XYZ for the period April-May 2024, with a total of 23,002 data points. The research stages included data selection, data pre-processing, normalization using Z-score, clustering with AHC, and evaluation using the Silhouette Coefficient and Davies-Bouldin Index (DBI). The results showed that the AHC method was able to form a representative grouping of machine conditions. Evaluation using the Silhouette Coefficient produced the best number of clusters at 2 clusters with a value of 0.591, indicating that the clustering quality was in the good category. Meanwhile, evaluation using DBI showed the best number of clusters at 4 clusters with a value of 0.404, indicating that the separation between clusters was quite good. These findings can be used as a reference in determining preventive machine maintenance policies so as to increase production efficiency.

KEYWORDS: Screw Press, Agglomerative Hierarchical Clustering, Clustering, Silhouette Coefficient, Davies-Bouldin Index

I. INTRODUCTION

In the palm oil industry, Fresh Fruit Bunches (FFB) are processed to produce crude palm oil (CPO) and palm kernel. The success of this process is largely determined by the performance of production machinery, one of which is the screw press machine that functions as a key component in the oil extraction stage from palm fruit [1]. Continuous operation of the machine over a long period of time can cause a decline in performance during certain periods. This condition not only has the potential to hamper the smooth running of the production processes, but also increases the risk of workplace accidents and causes significant losses for the company [2]. Therefore, it is important to ensure that production machines are always in optimal condition through the implementation of an effective maintenance system. Routine and periodic maintenance is a preventive measure to maintain stable machine performance and support production continuity [3]. Disruption to the mechanical components of the screw press machine will hamper the next processing stage and automatically cause losses [4].

A screw press machine is a tool that functions to continue the process of separating oil from the digester. This machine is equipped with a double screw that pushes the pressed mass out, while opposite pressure is applied through a hydraulic double cone. At this stage, the fruit pulp that has been stirred will be squeezed so that the oil contained in it can come out due to the pressing pressure. In a study with a clustering case study on screw press machines using the Fuzzy C-Means algorithm with evaluation using the Elbow method and Davies-Bouldin Index (DBI), it was found that the Elbow method produced four optimal clusters, while the DBI produced two optimal clusters [5]. Another study with a clustering case study on screw press machines using the K-Means algorithm with evaluation using the Elbow method and Davies-Bouldin Index (DBI) showed that the Elbow method produced four optimal clusters, while the DBI method produced three optimal clusters [6]. The screw press machine functions to squeeze chopped chips, crushed from the digester, to produce crude oil [7]. To improve oil separation efficiency, the pressings from the screw press are flowed through a Back Pressure Vessel (BPV), which functions to regulate back pressure so that oil extraction is maximized. The Back Pressure Vessel (BPV) is a pressurized vessel used to collect and distribute low-pressure steam to processing units in palm oil mills that utilize steam as a power source for equipment [8]. The main role of the BPV is to maintain a continuous supply of steam for various processes in palm oil mills. The steam is produced by boilers through the process of heating pressurized water, then part of it is used by turbines to generate electrical energy, while the remaining steam is channeled to the BPV. From the BPV, the steam is distributed to the stations that need it. Thus, the steam capacity produced by the boiler must be able to accommodate the needs of the entire processing chain in the palm oil mill

[9]. Given the crucial role of screw press machines and Back Pressure Vessels (BPV) in maintaining the smooth operation of the production process at palm oil mills, analyzing the operational data of these machines is essential. One approach that can be used to group this data is the clustering method.

Clustering is a method in data mining that falls under the unsupervised category. In general, clustering methods can be divided into two types, namely hierarchical clustering and non-hierarchical clustering, both of which are widely used in the data grouping process [10]. Clustering itself can be understood as the process of grouping data or objects based on the information contained in the data, which represents the characteristics of objects and their relationships [11]. The purpose of this process is so that objects included in one group have high similarity or relevance, while objects in different groups do not have such similarity or relevance [12]. The purpose of clustering is to identify or simplify data by dividing it into several small groups so that it is easier to analyze. This process is often referred to as data segmentation [13]. Clustering can also be used to reduce complex data by grouping similar data, thereby making the data shorter without losing its content [14]. Based on the various clustering methods available, the hierarchical approach is considered superior for certain cases because it is able to describe the relationship between data in the form of a hierarchical tree. One of the most commonly used methods is Agglomerative Hierarchical Clustering (AHC).

Agglomerative Hierarchical Clustering (AHC) is a clustering method that uses data analysis exploration techniques by grouping data into several groups called clusters [15]. AHC is a data grouping method based on the level of similarity between objects, resulting in a representation in the form of a tree-like hierarchical structure [16]. AHC uses several approaches to determine the distance between clusters, including single linkage, complete linkage, average linkage, and Ward [17].

In this study, the author used the Agglomerative Hierarchical Clustering (AHC) method, which has advantages over other clustering methods because it does not require determining the number of clusters at the beginning of the process. In addition, this study applied a single linkage approach, which is a method that determines the distance between clusters based on the closest proximity between objects in each cluster [18]. In a study with a case study of clustering to determine student majors using the AHC method, four clusters were produced, namely cluster 0 with 93 data points, cluster 1 with 10 data points, cluster 2 with 10 data points, and cluster 3 with 8 data points [19]. Another study with a case study of student literature clustering using the AHC method with a single linkage approach produced three clusters, namely cluster 1 with 1840 data, cluster 2 with 34 data, and cluster 3 with 16 data, with a total of 1890 data [20].

Based on the above explanation and issues, the purpose of this study is to obtain the best number of clusters using the Agglomerative Hierarchical Clustering method with original data from PT.XYZ from April 2024 to May 2024.

II. RESEARCH METHODOLOGY

A. Research Stages

This research consists of several stages that the author carried out systematically through a process of analysis and processing of relevant data. The first stage was data selection to determine which data would be used. Next, pre-processing was carried out by cleaning the data so that it was ready for transformation. After that, the data was transformed, which included a normalization process. The next stage was clustering using the Agglomerative Hierarchical Clustering (AHC) algorithm. The final stage was an evaluation to determine the best number of clusters produced by applying the algorithm. The research flow can be seen in Figure 1.

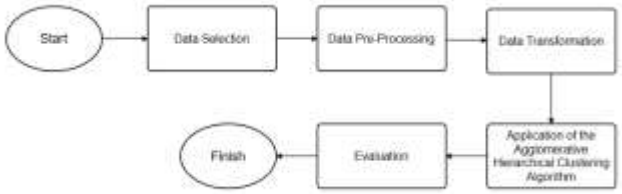

Figure 1. Research Methodology Flow

B. Data Selection

At this stage, data selection is carried out to determine which data will be processed so that it is in line with the research requirements. The variables and types of data used are adjusted to the research objectives so that the selected data are truly relevant. The purpose of this stage is to ensure that the data is ready for analysis so that it produces more accurate results.

C. Data Pre-Processing

One of the stages in data pre-processing is data cleaning. This stage aims to ensure that the selected and combined data are appropriate and ready for use in the processing stage. Data cleaning includes correcting invalid data, removing irrelevant attributes, and checking and handling missing values.

D. Data Transformation

The transformation stage aims to ensure that the data used can be optimally processed by the Agglomerative Hierarchical Clustering algorithm. At this stage, data normalization is carried out to maintain the validity and consistency of the data quality. Normalization is the process of transforming data into a more uniform, organized, and simple form for analysis, as well as reducing inconsistencies between values. This study uses the Z-score method. Z-score converts each data value into a standard distribution with a mean of 0 and a standard deviation of 1, so that data from various variables are on a comparable scale [21]. The following is the formula for using the Z-score.

$$Z = \frac{X - \mu}{\sigma} \quad (1)$$

Where X is the original data value, μ is the mean value, and σ is the standard deviation. By using this method, the data becomes more uniform on a comparable scale, so that differences in values between attributes do not dominate the distance calculations used in clustering.

E. Application of the Agglomerative Hierarchical Clustering Algorithm

Agglomerative Hierarchical Clustering (AHC) is a clustering technique that starts by treating each data point as its own cluster. The algorithm then iteratively merges the clusters that have the smallest distance between them, forming a hierarchical structure. One common method used in AHC is Single Linkage (Nearest Neighbor). In this approach, the distance between two clusters is determined based on the minimum distance between pairs of objects from different clusters. This approach is suitable when the purpose of the analysis is to group data based on the closest proximity between objects [22]. The application of the Agglomerative Hierarchical Clustering algorithm with the Single Linkage approach is as follows [23].

a. Calculate the distance between objects using Euclidean Distance :

$$d_{im} = \sqrt{\sum_{j=1}^{p}(x_{ij} - x_{mj})^2} \quad (2)$$

Explanation :

d_{im} = distance between object i and object m

x_{ij} = value of object i in variable j

x_{mj} = value of object m in variable j

p = number of variables

Then form the matrix $D = \{d_{im}\}$.

b. Determine the smallest distance in the matrix D and combine the two objects into one cluster.

c. Update the distance matrix using the Single Linkage approach :

$$d_{UV} = \min\{d_{UW}, d_{VW}\} \quad (3)$$

Explanation :

d_{UW} = distance between cluster U and cluster W

d_{VW} = distance between cluster V and cluster W

d. Continue until all objects are grouped into a single cluster.

F. Evaluation

In this study, evaluation was performed using the Silhouette Coefficient and Davies-Bouldin Index methods.

a. Silhouette Coefficient

The Silhouette Coefficient is one of the evaluation methods used to assess the quality of clustering results. This method measures how well an object is placed in the appropriate cluster compared to other clusters. The closer the Silhouette Coefficient value is to 1, the better the clustering quality within a cluster. Conversely, if the value is closer to -1, it indicates that the clustering quality in that cluster is poor or less than optimal [24].

$$s(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}} \quad (4)$$

Explanation :

$a(i)$ = average distance between object$i$ and all other objects in the same cluster

$b(i)$ = average distance between object$i$ and all objects in the nearest different cluster

$s(i)$ = Silhouette Coefficient value for object i

b. Davies-Bouldin Index

The Davies-Bouldin Index (DBI) is an evaluation method in clustering used to assess the quality of separation between clusters. The measurement is based on the level of cohesion within a cluster and the level of separation between clusters. A smaller DBI value indicates better clustering quality, as it indicates that the clusters formed are more compact and have a clear separation distance between one another [25]. The following are the steps for calculating the Davies-Bouldin Index (DBI).

1. Calculate the Sum of Squares Within-Cluster (SSW) value

$$SSW_i = \frac{1}{m_i}\sum_{j=i}^{m_i} d(x_j, c_i) \quad (5)$$

Explanation :

m_i = number of data points in the i-th cluster i

c_i = centroid of the cluster i

$d(x_j, c_i)$ = distance of each data point to centroid i calculated using Euclidean distance

2. Calculating the Sum of Squares Between-Cluster (SSB) value

The sum of squares between clusters (SSB) is used to assess the level of separation, which is calculated by measuring the distance between the cluster centroid and the overall data center. The higher the SSB value, the better the separation between clusters. The following is the SSB calculation formula.

$$SSB_i = (c, cj) \quad (6)$$

Where, $d(xi, xj)$ is the distance between data point$i$ and data point$j$ in another cluster.

3. Calculate the ratio of the Sum of Square Within-Cluster (SSW) value and the Sum of Square Between-Cluster (SSB)

A good cluster is one that has the smallest possible cohesion value and the largest possible separation value. This ratio is used to compare clusters, particularly between the$i$ and$j$ clusters, where the indices i ,j , and k represent the total number of clusters formed.

$$R_{ij} = \frac{SSW_i + SSW_i}{SSB_{ij}} \quad (7)$$

Explanation :

SSW_i = Sum of Square Within-Cluster at the centroid i

SSB_{ij} = Sum of Square Between-Cluster data between the$i$ and$j$ in different clusters

4. Calculate the Sum of Square Between-Cluster (DBI) value from the Ratio value obtained previously

$$DBI = \frac{1}{k}\sum_{i=1}^{k} max_{i\neq j}(R_{ij}) \quad (8)$$

Explanation :

Where k is the number of clusters used. The smaller the Davies-Bouldin Index (DBI) value obtained (non-negative, ≥ 0), the better the quality of the clusters formed. Conversely, if the DBI value obtained is quite large, this indicates that the clustering results using the clustering algorithm are suboptimal.

III.RESULT AND DISCUSSION

A. Data Selection

This study uses data obtained directly from PT.XYZ is a company engaged in palm oil processing. This study focuses on screw press machines, which play an important role in the palm oil extraction process. The data used comes from a screw press machine with the device code BPV EP01-09, which serves as a unique identifier for the machine in the company's monitoring system. A total of 23,002 data points were collected, covering various operational parameters such as machine code, pressure, and temperature. After the input process, only two attributes were selected for use, namely temperature and pressure, as these two attributes would be used for further processing. The following are the results of the data after the selected data was extracted, as shown in Table 1.

Table 1. Selected Data

NO	TEMPERATURE	PRESSURE
0	123	3,19
1	123	3
2	123	2,95
3	123	2,77
4	123	2,83
...	...	...
22997	0	2,99
22998	100	1,62
22999	100	1,53
23000	100	2,12
23001	100	2,03

After going through the selection stage, the data used in this study are presented in Table 1. The focus of the study was directed at two selected attributes, as they were considered the most relevant and had a significant influence. The selection of attributes considers data quality, contribution to target variables, and suitability to the context of the problem being studied.

B. Data Pre-Processing

This stage is a crucial step to ensure the quality and suitability of data before it is used in the analysis and modeling process. In this study, pre-processing focuses on data cleaning, which aims to remove or correct inconsistent data, missing values, and invalid data. The data cleaning process is carried out to improve data quality and make it more effective for use in the next stage of analysis. The data used was sourced from PT.XYZ in raw form. After selection based on device codes, the next stage was to check for possible duplications and missing values, as well as to ensure that the data to be analyzed was of a uniform and relevant type. With this checking and cleaning, the dataset became more ready and effective for use, resulting in more accurate clustering results. There are two attributes used, namely TEMPERATURE and PRESSURE, each of which has the same amount of data, namely 23,002 data points, no missing values (Not-Null), and a consistent data type, namely float64. These conditions indicate that the dataset has undergone a thorough data cleaning process and is ready for further analysis. With no missing values or data type differences, this clean dataset is ready to be used in the data transformation stage to improve the effectiveness of screw press machine condition modeling.

C. Data Transformation

Data transformation is necessary to address the difference in scale between the temperature (0–157) and pressure (0– 4.43) attributes. Without normalization, the attribute with the larger value range, namely temperature, will be more dominant in the distance calculation process in the Agglomerative Hierarchical Clustering (AHC) algorithm with the Single Linkage approach. This study uses the Z-score method. This transformation changes the value of each attribute based on the mean and standard deviation of that attribute, so that the data has a mean of 0 and a standard deviation of 1. The results of the transformed data can be seen in Table 2.

Table 2. Data After Transformation

NO	TEMPERATURE	PRESSURE
0	0,693	0,827
1	0,693	0,589
2	0,693	0,526
3	0,693	0,301
4	0,693	0,376
...	...	...
22997	-1,935	0,576
22998	0,202	-1,139
22999	0,202	-1,252
23000	0,202	-0,513
23001	0,202	-0,626

D. Application of the Agglomerative Hierarchical Clustering Algorithm

The clustering process in this study was performed using the Agglomerative Hierarchical Clustering algorithm on 23,002 data points that had undergone normalization and consisted of two main attributes, namely TEMPERATURE and PRESSURE. The purpose of applying this method was to form several clusters that represented similarities in the operational patterns of screw press machines. At each aggregation stage, the Agglomerative Hierarchical Clustering algorithm was applied using formulas (2) and (3) consistently on the normalized data (Table 2). In the clustering process, a threshold value is used to determine the maximum distance between clusters to be merged. This threshold value serves as a cut-off distance on the dendrogram, which is the point at which the cluster merging process is stopped. With a threshold, the system can automatically form the number of clusters that matches the existing data structure without the need to determine the number of clusters at the outset. The selection of the appropriate threshold affects the final grouping results. The smaller the threshold value, the more clusters are formed, and conversely, the larger the threshold value, the fewer clusters are produced. Through this approach, each data point obtains a label according to the cluster to which it belongs, so that the grouping results are able to describe the condition of the machine in a more structured group. The information from this clustering can then be used as a basis for further analysis, such as identifying abnormal conditions or detecting potential machine damage. The test results can be seen in Table 3.

Table 3. Test Result with 4 Clusters

CLUSTER	NUMBER OF DATA
C0	18529
C1	4396
C2	69
C3	8

Based on Table 3, the results of the data after testing using the Agglomerative Hierarchical Clustering algorithm show the number of data in each cluster as shown in Table 5. The number of data in C0 is 18,529, the number of data in C1 is 4,396, the number of data in C2 is 69, and the number of data in C3 is 8. The next step is visualization. The visualization can be seen in Figure 2.

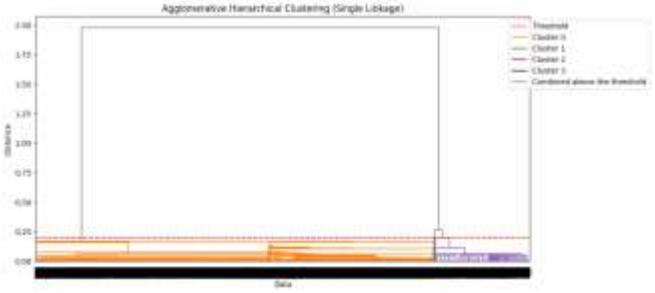

Figure 2. Visualization of clustering results

E. Evaluation

After applying the Agglomerative Hierarchical Clustering algorithm with various cluster numbers, the next step is to evaluate and interpret the quality of the clustering results. The purpose of this step is to assess the extent to which the data has been successfully grouped according to its characteristics, while ensuring that the number of clusters selected is truly optimal. In this study, two evaluation methods were used, namely the Silhouette Coefficient and the Davies-Bouldin Index (DBI). These two methods complement each other in determining the number of clusters that are most representative of the data structure, so that the clustering results obtained are not only technically accurate but also relevant in the context of analyzing the condition of screw press machines.

a. Silhouette Coefficient

Evaluating clustering results is very important to ensure that the division of data into clusters is optimal. The Silhouette Coefficient method can be used to assess the quality of each cluster by looking at the resulting coefficient value. The closer the value is to 1, the better the clustering quality because it indicates that the data is more similar to the cluster it belongs to than to other clusters. Conversely, a value close to 0 indicates overlap between clusters. The following are the results of the Silhouette Coefficient based on the number of clusters tested. The results of the evaluation using the silhouette coefficient can be seen in Table 4.

Table 4. Evaluation of Cluster Result Using the Silhouette Coefficient

CLUSTER	SILHOUETTE VALUE
2	0,591
3	0,565
4	0,553
5	-0,176
6	-0,287
7	-0,153
8	-0,169
9	-0,183
10	-0,184

Based on Table 4, the clustering results were evaluated using the Silhouette Coefficient method to assess the quality of data separation in each cluster. The closer the value is to 1, the better the clustering quality because it indicates a clearer distance between clusters. The Silhouette Coefficient was calculated using formula (4) on the data shown in Table 2, with the calculation results presented in Table 4. Based on these results, the highest Silhouette value was obtained for two clusters (k = 2) with a score of 0.591. This indicates that the configuration with four clusters provides the most optimal separation and density. To clarify the evaluation trend, the calculation results are visualized in the form of a dendrogram so that the pattern of changes in the Silhouette value against

variations in the number of clusters can be observed more intuitively and support the determination of the most appropriate number of clusters. The following is the evaluation result with the Silhouette Coefficient as shown in Figure 3.

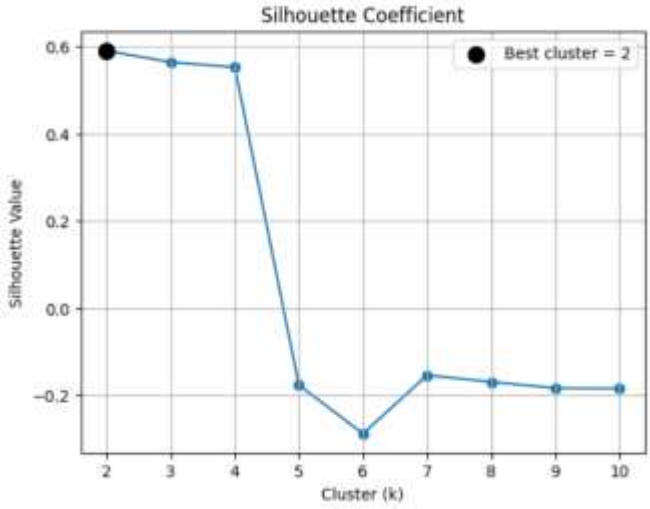

Figure 3. Evaluation Graph Using Silhouette Coefficient

b. Davies-Bouldin Index (DBI)

Evaluating clustering results plays an important role in ensuring that data division into clusters is optimal. Using the Davies-Bouldin Index (DBI) method, the quality of each cluster can be measured by selecting the number of clusters that produce the smallest DBI value. The lower the DBI value obtained, the better the quality of the grouping. The following shows the results of DBI value calculations based on the number of clusters tested. The results of the evaluation using DBI can be seen in Table 5.

Table 5. Evaluation of Cluster Results Using DBI

CLUSTER	DBI VALUE
2	0,586
3	0,453
4	0,404
5	1,394
6	1,523
7	1,667
8	1,571
9	1,527
10	1,453

Based on Table 5, the evaluation of clustering results using the Davies-Bouldin Index (DBI) method aims to assess the quality of data division into each cluster, where the smaller the DBI value obtained, the better the clustering quality. The DBI calculation was performed using formulas (5) to (8) on the data in Table 2, resulting in the DBI values shown in Table 5. From these calculations, it can be seen that the lowest DBI value was achieved with three clusters (k = 4) with a value of 0.404. This indicates that the configuration with four clusters provides the most optimal separation and density. To provide a clearer picture of the evaluation results trend, a visualization in the form of a dendrogram was also carried out, so that the pattern of changes in DBI values against variations in the number of clusters can be observed more intuitively and support decision-making regarding the most appropriate number of clusters.

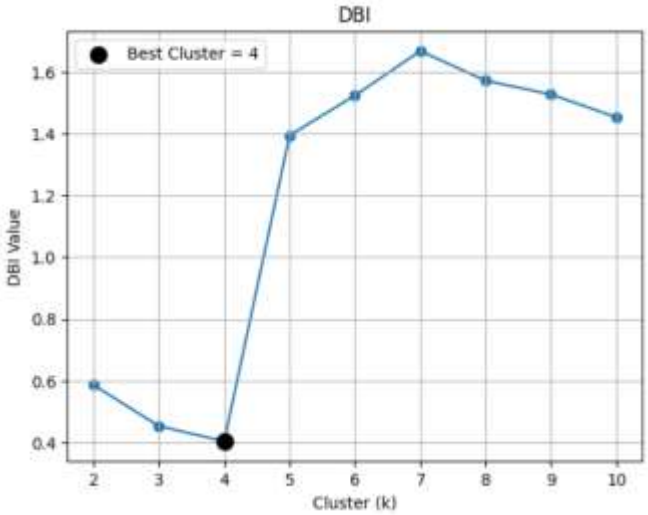

Figure 4. Evaluation Graph Using DBI

F. Comparison of the Best Cluster Results

A comparison of clustering results using the Silhouette Coefficient and DBI methods was conducted to determine the most appropriate number of clusters, taking into account the characteristics of each cluster formed. The best cluster results obtained from the Silhouette Coefficient method were 2 clusters, while the best cluster results from the DBI method were 4 clusters. The following are the results of the 2-cluster and 4-cluster tests.

Table 6. Testing of 2 clusters

CLUSTER	NUMBER OF DATA	TEMPERATURE	PRESSURE
C0	4473	0	0,02-3,32
C1	18529	92,5-157	0-4,43

Table 7. Testing of 4 clusters

CLUSTER	NUMBER OF DATA	AVERAGE TEMPERATURE	AVERAGE PRESSURE
C0	18529	92,5-157	0-4,43
C1	4396	0	0,77-3,32
C2	69	0	0,02-0,23
C3	8	0	0,44-0,61

Based on Table 6 and Table 7, the results of clustering evaluation using two methods, namely Silhouette Coefficient and Davies-Bouldin Index (DBI), obtained different results regarding the best number of clusters. The Silhouette Coefficient method indicates that the optimal number of clusters is 2, because at this number the Silhouette value is

higher, indicating clearer data separation, compactness, and a good level of similarity within clusters. Meanwhile, the Davies-Bouldin Index (DBI) method shows the best results with 4 clusters, because the DBI value produced is lower, which means that the level of similarity between clusters is smaller and the distance between clusters is relatively better. In this study, the use of the Silhouette Coefficient method with the formation of 2 clusters is considered more appropriate because it is able to provide a more concise, structured data separation that is easier to understand in the analysis process.

IV. CONCLUSION

This study discusses the application of the Agglomerative Hierarchical Clustering (AHC) algorithm with a single-linkage approach in grouping screw press machine conditions based on two main attributes, namely temperature and pressure. The data used is original data from PT. XYZ for the period April - May 2024 with a total of 23,002 data points. The research process was carried out in stages through data selection, pre-processing with data cleaning, data transformation using Z-Score normalization, application of the AHC algorithm, and evaluation of clustering results using the Silhouette Coefficient and Davies-Bouldin Index (DBI) methods. The results showed that the clustering process divided the data into several groups that reflected the conditions of the screw press machine in different situations. Based on the evaluation using the Silhouette Coefficient, the best number of clusters was obtained in two clusters with a value of 0.591. This indicates that the two-cluster configuration is able to provide clearer, more compact, and structured data separation. Meanwhile, the evaluation results using the Davies-Bouldin Index (DBI) showed that the best number of clusters was four clusters with a DBI value of 0.404, which indicates that the quality of separation between clusters is quite good. The difference in these evaluation results indicates a difference in focus between the visual clarity of clustering and the level of separation between clusters. Overall, this study proves that the AHC algorithm with the Single Linkage approach can be used to analyze the condition of screw press machines in a more systematic and structured manner. These clustering results can be used as a basis for determining the number of clusters that are most representative of data patterns. The limitation of this study is that it only uses two main attributes, so for further research, it is recommended to add other attributes in order to obtain more comprehensive and accurate clustering results.

REFERENCES

1. Y. Ningsih, M. Khudari, T. Linangsari, M. Zein, and E. Lestari, "Evaluasi Kinerja Mesin Screw Press Melalui Penerapan Total Productive Maintenance (TPM) di Pabrik Kelapa Sawit PT. XYZ," *J. Teknol. Agro-Industri*, vol. 12, no. 1, pp. 61–69, 2025, doi: 10.34128/jtai.v12i1.234.
2. M. L. Volodymyr Havran, "Determination of Hopper Fullness of Smart Screw Press Using Machine Learning," *Comput. Des. Syst. Theory Pract.*, vol. 6, no. 1, pp. 161–168, 2024, doi: 10.23939/cds2024.01.161.
3. F. Pohan, I. Saputra, and R. Tua, "Scheduling Preventive Maintenance to Determine Maintenance Actions on Screw Press Machine," *J. Ris. Ilmu Tek.*, vol. 1, no. 1, pp. 1–12, 2023, doi: 10.59976/jurit.v1i1.4.
4. D. Wardianto and Anrinal, "Analisis kegagalan mesin screw press failure analysis of the screw press machine," *J. Tek. Mesin Insitut Teknol. Padang*, vol. 12, no. 1, pp. 2089–4880, 2022.
5. J. Jasril, M. F. Al Fiqri, S. Sanjaya, L. Handayani, and F. Insani, "Pengelompokan Data Kondisi Mesin Screw Press Menggunakan Algoritma Fuzzy C-Means," *Inf. Syst. J.*, vol. 8, no. 01, pp. 60–70, 2025, doi: 10.24076/infosjournal.2025v8i01.2133.
6. F. K. Rahman, J. S. Sanjaya, L. Handayani, and F. Insani, "Penerapan Algoritma K-Means Clustering pada Kinerja Mesin Screw press," *Bull. Inf. Technol.*, vol. 6, no. 2, pp. 59–70, 2025, doi: 10.47065/bit.v5i2.1783.
7. M. I. Pasaribu, D. A. A. Ritonga, and A. Irwan, "Analisis Perawatan (Maintenance) Mesin Screw Press Di Pabrik Kelapa Sawit Dengan Metode Failure Mode and Effect Analysis (Fmea) Di Pt. Xyz," *Jitekh*, vol. 9, no. 2, pp. 104–110, 2021, doi: 10.35447/jitekh.v9i2.432.
8. Z. Effendi, I. U. P. Rangkuti, and M. D. Nugroho, "Analisa Pengaruh Laju Massa Uap Rata-Rata Terhadap Kualitas Uap Rata-Rata Pada Pipa BPV (Back Pressure Vessel) Ke Sterilizer Di Pabrik Kelapa Sawit Kapasitas 50 Ton / Jam Dalam berkembangnya industri minyak kelapa sawit yang mengalami kemajuan sangat pe," vol. 13, no. 01, pp. 148–155, 2024.
9. M. Wisnu, G. Supriyanto, and Hermantoro, "Analisis Kebutuhan Uap pada Perebusan Tiga Puncak," *Agroforetech*, vol. 1, pp. 685–692, 2023.
10. M. Norshahlan, H. Jaya, and R. Kustini, "Penerapan Metode Clustering Dengan Algoritma K-means Pada Pengelompokan Data Calon Siswa Baru," *J. Sist. Inf. Triguna Dharma (JURSI TGD)*, vol. 2, no. 6, p. 1042, 2023, doi: 10.53513/jursi.v2i6.9148.
11. S. D. K. Wardani, A. S. Ariyanto, M. Umroh, and D. Rolliawati, "Perbandingan Hasil Metode Clustering K-Means, Db Scanner & Hierarchical Untuk Analisa Segmentasi Pasar," *JIKO (Jurnal Inform. dan Komputer)*, vol. 7, no. 2, p. 191, 2023, doi: 10.26798/jiko.v7i2.796.
12. R. A. Ananda, "Clustering Menggunakan Algoritma K-Means untuk Mengelompokan Data Perjudian Berdasarkan Wilayah di Kota Binjai (Studi Kasus :

Pengadilan Negeri Binjai)," no. 4, 2024.
13. S. Maghfiroh, A. P. Wijaya, and A. Hidayat, "Implementasi Clustering Pada Data Metering Mesin Di Stasiun Transmisi Trans Tv Semarang," *Pros. Sains Nas. dan Teknol.*, vol. 13, no. 1, p. 161, 2023, doi: 10.36499/psnst.v13i1.9078.
14. P. Tang, H. Tang, W. Wang, and Y. Liu, "Contrastive Clustering," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 13833 LNCS, pp. 294–305, 2023, doi: 10.1007/978-3-031-27077-2_23.
15. Ulfatul Syahara, Esty Kurniawati, Mario Putra Suhana, Rika Anggraini, and Falmi Yandri, "Penerapan Metode AHC (Agglomerative Hierarchical Clustering) untuk Klasifikasi Habitat Bentik di Desa pengudang, Kabupaten Bintan," *INSOLOGI J. Sains dan Teknol.*, vol. 3, no. 3, pp. 306–314, 2024, doi: 10.55123/insologi.v3i3.3547.
16. K. Pratama Simanjuntak and U. Khaira, "Hotspot Clustering in Jambi Province Using Agglomerative Hierarchical Clustering Algorithm," *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 1, no. 1, pp. 7–16, 2021.
17. R. P. Justitia, N. Hidayat, and E. Santoso, "Implementasi Metode Agglomerative Hierarchical Clustering Pada Segmentasi Pelanggan Barbershop (Studi Kasus : RichDjoe Barbershop Malang)," *J. Pengemb. Teknol. Inf. Dan Ilmu Komput.*, vol. 5, no. 3, pp. 1048–1054, 2021, [Online]. Available: https://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/8730
18. R. Kusumastuti, E. Bayunanda, A. M. Rifa'i, M. R. G. Asgar, F. I. Ilmawati, and K. Kusrini, "Clustering Titik Panas Menggunakan Algoritma Agglomerative Hierarchical Clustering (AHC)," *CogITo Smart J.*, vol. 8, no. 2, pp. 501–513, 2022, doi: 10.31154/cogito.v8i2.438.501-513.
19. R. T. Aldisa, "Data Mining Penentuan Jurusan Siswa Menggunakan Metode Agglomerative Hierarchical Clustering (AHC," *J. Media Inform. Budidarma*, vol. 7, no. 2, p. 873, 2023, doi: 10.30865/mib.v7i2.6092.
20. A. N. Fadhilah and A. Jananto, "Klasterisasi Literatur Mahasiswa Menggunakan Metode Ahc Di Dinas Kearsipan Dan Perpustakaan Provinsi Jawa Tengah," *J. Din. Inform.*, vol. 13, no. 1, pp. 09–17, 2021, doi: 10.35315/informatika.v13i1.8366.
21. S. E. Saqila, I. P. Ferina, and A. Iskandar, "Analisis Perbandingan Kinerja Clustering Data Mining Untuk Normalisasi Dataset," *J. Sist. Komput. dan Inform.*, vol. 5, no. 2, p. 356, 2023, doi: 10.30865/json.v5i2.6919.
22. A. P. Wijaya, A. S. Kinanthi, D. Yanawati, and S. Syamsidar, "Pengelompokkan Kabupaten / Kota di Pulau Jawa Berdasarkan Faktor Kemiskinan Menggunakan Metode Hierarchical Clustering," *J. Sains dan Manaj.*, vol. 13, no. 1, pp. 40–51, 2025.
23. C. B. G. Allo, "Clustering Kabupaten/Kota di Provinsi Papua Berdasarkan Produk Domestik Regional Bruto Menurut Lapangan Usaha Menggunakan Single Linkage dan K-Medoids," *J. Gaussian*, vol. 13, no. 1, pp. 111–120, 2024, doi: 10.14710/j.gauss.13.1.111-120.
24. T. Rahmawati, Y. Wilandari, and P. Kartikasari, "Analisis Perbandingan Silhouette Coefficient Dan Metode Elbow Pada Pengelompokkan Provinsi Di Indonesia Berdasarkan Indikator Ipm Dengan K-Medoids," *J. Gaussian*, vol. 13, no. 1, pp. 13–24, 2024, doi: 10.14710/j.gauss.13.1.13-24.
25. S. G. Ahmad, D. Arifianto, and W. Suharso, "Jurnal Smart Teknologi Jurnal Smart Teknologi," *Univ. Muhammadiyah Jember*, vol. 3, no. 5, pp. 502–510, 2022.