Journal of Artificial Intelligence and Engineering Applications

Website: https://ioinformatic.org/

15th February 2025. Vol. 4. No. 2; e-ISSN: 2808-4519

Use of K-Means Algorithm in Model Improvement Production Data Grouping for Determination Convection Production Strategy

Ica Pandia[1*], Martanto[2], Arif Rinaldi Dikananda[3], Dede Rohman[4]

[1,2,3,4]*STMIK IKMI Cirebon*
isnaisna236@gmail.com[1]*

Abstract

This research was conducted to implement the K-Means Clustering algorithm in grouping convection production data to support the improvement of efficiency and effectiveness of production strategies. The data used is PT ABC's production data, which consists of important attributes, such as Production ID, Production Date, Product Name, Product Type, Color, Size, Raw Materials, and Order. The research method follows the stages of Knowledge Discovery in Database (KDD), which includes selection, preprocessing, transformation, data mining, and evaluation, so as to ensure that the data processed is relevant and ready to be analyzed. The grouping process is carried out using the K-Means algorithm, which groups data based on attribute similarity by determining the optimal number of clusters. The evaluation of the clustering results was carried out using the Silhouette Score and Davies-Bouldin Index metrics, where the results showed values that represented good cluster quality. A high Silhouette Score indicates that the data in the cluster has good uniformity, while a low Davies-Bouldin Index indicates a clear distance between clusters. The results of the grouping produce three main clusters that illustrate different production patterns, such as clusters with high, medium, and low order quantities. This analysis of the cluster provides important insights in supporting strategic decision-making, such as prioritizing resource management in high-order clusters and evaluating production efficiency in low-order clusters. This research is in line with previous literature that shows that the K-Means algorithm can be used effectively in big data grouping to support strategic planning. The practical contribution of this research is to help convection companies in understanding production patterns, so that production strategies can be designed more efficiently, responsively, and directionally. For further research, it is recommended to add new variables, such as production costs or work duration, as well as test other clustering algorithms to obtain more comprehensive results.

Keywords: K-Means, clustering, production data, production efficiency, convection strategy

1. Introduction

The rapid development of information technology in recent decades has changed many aspects of human life, including business, education, and industry. Digitalization and automation have made it possible for industries to collect, manage, and analyze data on an unprecedented scale. One of the important impacts of this development is the increased use of big data technology and data analysis for better and faster decision-making. In the manufacturing industry, especially convection, the use of data to analyze production processes and predict future needs is becoming increasingly important. Data mining technology, including clustering techniques, has been widely used to help companies understand complex data patterns and improve operational efficiency. One of the most widely used algorithms in clustering is K-Means, which allows grouping data based on a significant number of attribute similarities.
Various previous studies have discussed the application of the K-Means algorithm in the context of grouping production data. According to [1] in his research entitled "Analysis of the K-Means Algorithm on the Spreading Report of the Cutting Department of PT. Busana Indah Global" This research focuses on processing data spreading reports in the Cutting Department of PT. Busana Indah Global uses the k-means clustering algorithm to group fabric cutting data. This method includes quantitative data analysis from the H&M Mapple HW Flare Leggings Season 3 Spreading Report. The results show that the k-means clustering algorithm is able to group the data into three categories: the highest, medium, and lowest spreading, with details of 1 fabric cutting in the highest category, 13 in the medium, and 7 in the lowest. These findings help the cutting department in the production process and facilitate reporting when needed by the buyer or audit. Furthermore, according to [2] in his research entitled "Analysis of Bank Customer Segmentation Based on Credit Extraction Using the K-Means Clustering Method", this research focuses on segmenting bank customers using the K-means clustering algorithm to optimize financing offers. By segmenting customers based on credit behavior, this study aims to improve the effectiveness of targeting. The process includes data cleaning, transformation, grouping with K-means, and analysis of results. The results show that with 4 clusters, this method helps identify potential customers for a more targeted financing strategy. Furthermore, according to [3] entitled Data Clustering Using the K-Means Method to Recommend Academic Learning for Active Students in Extracurriculars. using the K-Means algorithm to segment the academic grades of grade VIII and IX students who are active in extracurriculars, with the aim of helping to balance their academic

achievement. The results showed that most students had high grades, with eight cluster categories, including smart, moderate, and adequate. These findings also suggest that students with moderate grades are more active in extracurriculars, while students who have a balance of science and social studies scores can choose majors according to their interests.

2. Literature Review

2.1. Related Research Results

Paper 1 according to [4] entitled *Clustering of Vaccine Priority Areas Using the K-Means Clustering Algorithm*, This study aims to develop a Covid-19 classification application in North Sumatra Province using the K-Means Clustering algorithm to determine vaccination priorities based on Covid-19 case data. The research methodology includes the stages of collecting population data that is eligible for vaccination and Covid-19 case data, designing the system interface, implementing the application, and testing to ensure the accuracy of the results. This application successfully grouped regions into three clusters: C1 (High), C2 (Medium), and C3 (Low). The results of clustering show that Medan City is included in the high cluster and prioritized in the vaccination program, while Deli Serdang is in the medium cluster, as well as most other areas in the low cluster.

Paper 3 according to [5] Titled *K-Means Clustering Algorithm in Predicting Recipients of Direct Cash Assistance (BLT) for Village Funds*, This research aims to overcome the problem in determining the right recipients of Village Fund Direct Cash Assistance (BLT-DD) in Nagari Taluk, Lintau Buo District, Tanah Datar Regency. The method used is the K-Means Clustering algorithm, which involves several stages, including problem identification and analysis, data collection, literature study, and calculation using the algorithm. This research was led by Yosep Filki, with the title "K-Means Clustering Algorithm in Predicting Recipients of Direct Cash Assistance (BLT) for Village Funds."

Paper 4 according to [6] entitled Image Searching Data Color Image Clustering Algorithm K Means Clustering Algorithm on Sales Data, This study analyzes the use of the K-Means Clustering algorithm in searching for sales data through color images, with the aim of helping users who have difficulty finding product names or codes using product images. The research process includes data collection to create an image-based sales dataset, user interface design, program development using the Scilab programming language, implementation of image search, and application effectiveness testing. The results show that the K-Means Clustering algorithm successfully clusters sales data with an accuracy rate of 93.33%, proving its effectiveness in supporting image-based product search.

Paper 8 by [7] entitled *Research on Vivo Smartphone Sales Grouping Using the K-Means Method.* This research aims to find patterns and strategies that can increase sales and marketing of Vivo products by utilizing transaction data. The method used is K-Means, which involves grouping sales data into low, medium, and high categories based on the minimum value of the group. The results show that K-Means can segment sales with 100% accuracy compared to manual methods, thus providing an overview of best-selling and in-demand products to support marketing decisions and decisions.

Paper 10 according to [8] entitled *Pixel clustering of images of the Monpera perverted photo collection with the K-Means method in augmented reality applications.* This research focuses on the development of an Augmented Reality application to introduce national heroes through a collection of photos at the Monpera Museum Palembang, using the K-Means method for pixel clustering of images. The method used is the Rational Unified Process (RUP), which consists of the inception stage for software requirements analysis, elaboration for comprehensive design, and testing the K-Means method on image pixel clustering. The results show that this application is effective for visualization and learning of the history of national heroes, as well as improving the image quality of the museum's photo collection.

Paper 11 according to [9] entitled *Clustering of farmer group coaching to increase corn production using the K-Means algorithm.* This research aims to group farmer groups to increase corn production by applying the K-Means algorithm. The process involves several stages, such as describing the problem, analyzing, determining the objectives, studying the literature, collecting data, and analyzing data processing techniques. The results showed that 56 farmer groups were in clusters with low production, 17 groups with medium production, and 7 groups with high production, providing guidance for more targeted coaching.

Paper 12 according to [10] entitled *Application of K-Means Clustering Method Data Mining for Sales Analysis at Banten Hijab Fashion Stores.* This study uses the K-Means clustering method to analyze sales data at the hijab fashion store, Toko Strand, with the aim of supporting a more effective marketing strategy. The process involves using the initial centroid, iterating until the data is stable in the cluster, and calculating the new centroid based on the cluster average. The results showed that 11 items were very hot-selling, 55 items were selling well, and 34 items were not in demand, providing insights to optimize sales management.

3. Research Methods

Methodology In this study, the Knowledge Discovery in Database (KDD) method is used. Below is a picture of this research flow method.

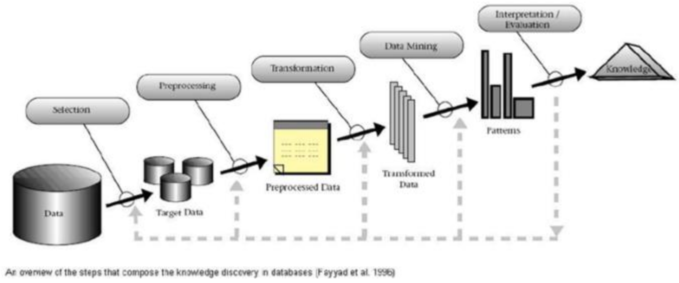


Figure 1: Stages of KDD Method Research

4. Results and Discussion

4.1. Data Selection

Table 1 below shows the Production Dataset consisting of 8 main attributes, namely Production ID, Production Date, Product Name, Product Type, Color, Size, Raw Material, and Order, which includes unique identification, product characteristics, production time, material requirements, and order quantities to support the analysis of production data clustering using the K-Means algorithm.

Table 1: Dataset

No	Id produksi	Tanggal produksi	Nama produk	Jenis produk	warna	ukuran	Bahan baku	order
1	c4	04/10/24	kemeja	kemeja kerja	zodiac night	s	cotton poplin	1904
2	c4	04/10/24	kemeja	kemeja kerja	zodiac night	m	cotton poplin	4128
3	c4	04/10/24	kemeja	kemeja kerja	zodiac night	l	cotton poplin	5000
4	c4	04/10/24	kemeja	kemeja kerja	zodiac night	xl	cotton poplin	3976
5	c4	04/10/24	kemeja	kemeja kerja	zodiac night	xxl	cotton poplin	2248
6	c4	04/10/24	kemeja	kemeja kerja	dog bone	s	cotton poplin	2564
7	c4	04/10/24	kemeja	kemeja kerja	dog bone	m	cotton poplin	5580
8	c4	04/10/24	kemeja	kemeja kerja	dog bone	l	cotton poplin	6740
8	c4	04/10/24	kemeja	kemeja kerja	dog bone	xl	cotton poplin	5344
…	…	………..	………	…………….	……………	……	…………….	……
251	c1brpll	04/10/24	dress	dress kasual	navy blue	xxl	polycotton belt	2802

4.2. Pre-Processing

Figure 2 Below shows the results of using Excel's Read Operator, which is used to import production datasets from Excel-formatted files into a data analysis platform. This operator functions to read data directly from external sources, so that the data can be further processed in the analysis stage. The results from this operator display a production dataset consisting of various attributes, such as Production ID, Production Date, Product Name, Product Type, Color, Size, Raw Material, and Order. The resulting output ensures that all the necessary data has been imported correctly and in a format appropriate for the pre-processing stage. With well-organized data, the analysis process becomes more efficient and accurate in producing optimal clustering results.

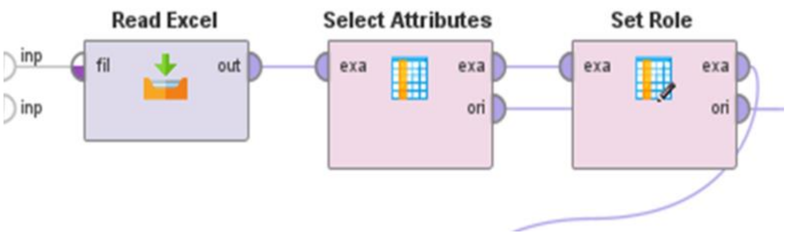
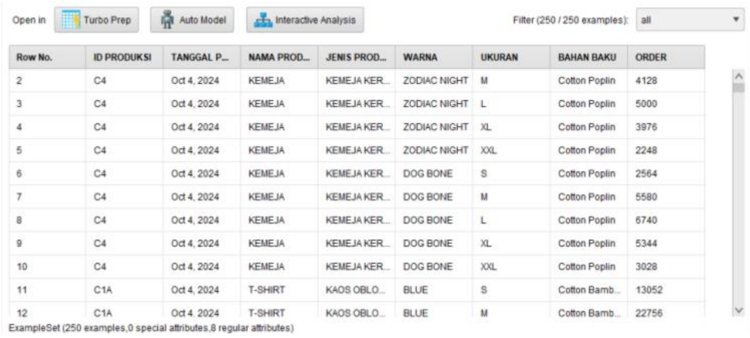


Figure 2: Showing the results of the operator

1. Operator Read Excel

Figure 3 below shows the Output of the Excel Read Operator, which is used to read and display production data from Excel files into the analysis platform. This output ensures that the imported dataset includes all the necessary attributes, such as Production ID, Production Date, Product Name, Product Type, Color, Size, Raw Material, and Order, with the appropriate data structure. The results displayed show the data in an organized table, which simplifies the process of verifying the data before moving on to the next stage of analysis. This view ensures that all data has been imported correctly without any information loss, thus providing a solid basis for pre-processing, transformation, and analysis steps using the K-Means algorithm. With validation through these outputs, input errors can be minimized, and analysis can run smoothly.



Row No.	ID PRODUKSI	TANGGAL P...	NAMA PROD...	JENIS PROD...	WARNA	UKURAN	BAHAN BAKU	ORDER
2	C4	Oct 4, 2024	KEMEJA	KEMEJA KER...	ZODIAC NIGHT	M	Cotton Poplin	4128
3	C4	Oct 4, 2024	KEMEJA	KEMEJA KER...	ZODIAC NIGHT	L	Cotton Poplin	5000
4	C4	Oct 4, 2024	KEMEJA	KEMEJA KER...	ZODIAC NIGHT	XL	Cotton Poplin	3976
5	C4	Oct 4, 2024	KEMEJA	KEMEJA KER...	ZODIAC NIGHT	XXL	Cotton Poplin	2248
6	C4	Oct 4, 2024	KEMEJA	KEMEJA KER...	DOG BONE	S	Cotton Poplin	2564
7	C4	Oct 4, 2024	KEMEJA	KEMEJA KER...	DOG BONE	M	Cotton Poplin	5580
8	C4	Oct 4, 2024	KEMEJA	KEMEJA KER...	DOG BONE	L	Cotton Poplin	6740
9	C4	Oct 4, 2024	KEMEJA	KEMEJA KER...	DOG BONE	XL	Cotton Poplin	5344
10	C4	Oct 4, 2024	KEMEJA	KEMEJA KER...	DOG BONE	XXL	Cotton Poplin	3028
11	C1A	Oct 4, 2024	T-SHIRT	KAOS OBLO...	BLUE	S	Cotton Bamb...	13052
12	C1A	Oct 4, 2024	T-SHIRT	KAOS OBLO...	BLUE	M	Cotton Bamb...	22756

ExampleSet (250 examples,0 special attributes,8 regular attributes)

Figure 3: EXCEL Read Output Operator

Table 2 below shows the data types of the production dataset that include eight main attributes, namely Production ID (polynominal), Production Date (date), Product Name, Product Type, Color, Size, Raw Material (polynominal), and Order (integer), each of which is designed to support the pre-processing process and clustering analysis using the K-Means algorithm.

Table 2: Production Dataset Type

No	nama	type data
1	id produksi	polynominal
2	tanggal produksi	date
3	nama produk	polynominal
4	jenis produk	polynominal
5	warna	polynominal
6	ukuran	polynominal
7	bahan baku	polynominal
8	order	integer

2. Operator Select Attributes

Figure 4 shows the Attribute Selection Process using the Select Attributes operator, which selects relevant attributes from the dataset to be used in the analysis. In this study, the selected attributes include Production ID, Production Date, Product Name, Product Type, Color, Size, Raw Material, and Order. This process aims to filter out attributes that are not relevant or unnecessary in the clustering analysis, so that only data that supports the research objectives will be further processed. By simplifying the dataset through this attribute selection, the analysis becomes more focused, efficient, and reduces the potential for interference from irrelevant attributes, so that the results of data grouping using the K-Means algorithm can be more accurate and representative.

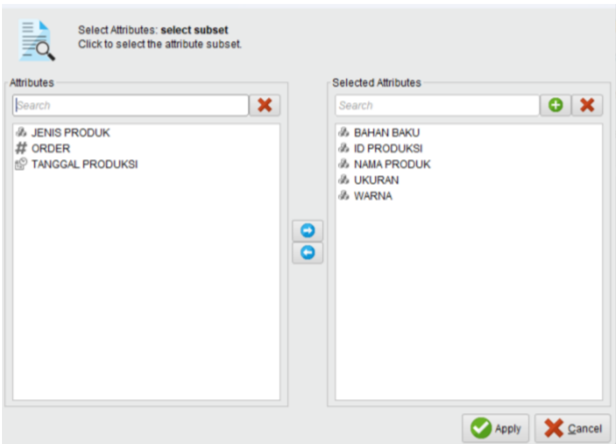


Figure 4: Attribute selection process

Figure 5 Below shows the Output of the Select Attributes Operator, which is used to display the results of the attribute selection of the production dataset. These operators ensure that only the attributes relevant for analysis, such as Production ID, Production Date, Product Name, Product Type, Color, Size, Raw Material, and Order, are retained. The results displayed are in the form of organized tables, which reflect the dataset with attributes that have been filtered according to the needs of the analysis. This process ensures that the dataset becomes more focused, reduces data complexity, and makes it easier for the K-Means algorithm to group data based on attributes that have a direct contribution to production patterns. With this output, the next step in the analysis can be carried out efficiently and accurately.

Row No.	ID PRODUKSI	NAMA PROD...	WARNA	UKURAN	BAHAN BAKU
1	C4	KEMEJA	ZODIAC NIGHT	S	Cotton Poplin
2	C4	KEMEJA	ZODIAC NIGHT	M	Cotton Poplin
3	C4	KEMEJA	ZODIAC NIGHT	L	Cotton Poplin
4	C4	KEMEJA	ZODIAC NIGHT	XL	Cotton Poplin
5	C4	KEMEJA	ZODIAC NIGHT	XXL	Cotton Poplin
6	C4	KEMEJA	DOG BONE	S	Cotton Poplin
7	C4	KEMEJA	DOG BONE	M	Cotton Poplin
8	C4	KEMEJA	DOG BONE	L	Cotton Poplin
9	C4	KEMEJA	DOG BONE	XL	Cotton Poplin
10	C4	KEMEJA	DOG BONE	XXL	Cotton Poplin
11	C1A	T-SHIRT	BLUE	S	Cotton Bamb...
12	C1A	T-SHIRT	BLUE	M	Cotton Bamb...
13	C1A	T-SHIRT	BLUE	L	Cotton Bamb...
14	C1A	T-SHIRT	BLUE	XL	Cotton Bamb...
15	C1A	T-SHIRT	BLUE	XXL	Cotton Bamb...

Figure 5: Output Operator Select Atribut

3. Operator Set Role

Figure 6 below shows the Set Role Process using the Set Role operator, which serves to define the role of each attribute in the production dataset. In this process, attributes such as Production ID are assigned as role ids for unique identification of data, while other attributes such as Production Date, Product Name, Product Type, Color, Size, Raw Material, and Order are assigned as regular roles to use in clustering analysis. This process ensures that the relevant attributes can be used optimally by the K-Means algorithm, while unnecessary attributes are ignored to avoid interference in the analysis. By assigning the role of attributes precisely, the dataset becomes more structured and tailored to the needs of the analysis, so that the algorithm can work efficiently to produce accurate and meaningful data groupings.

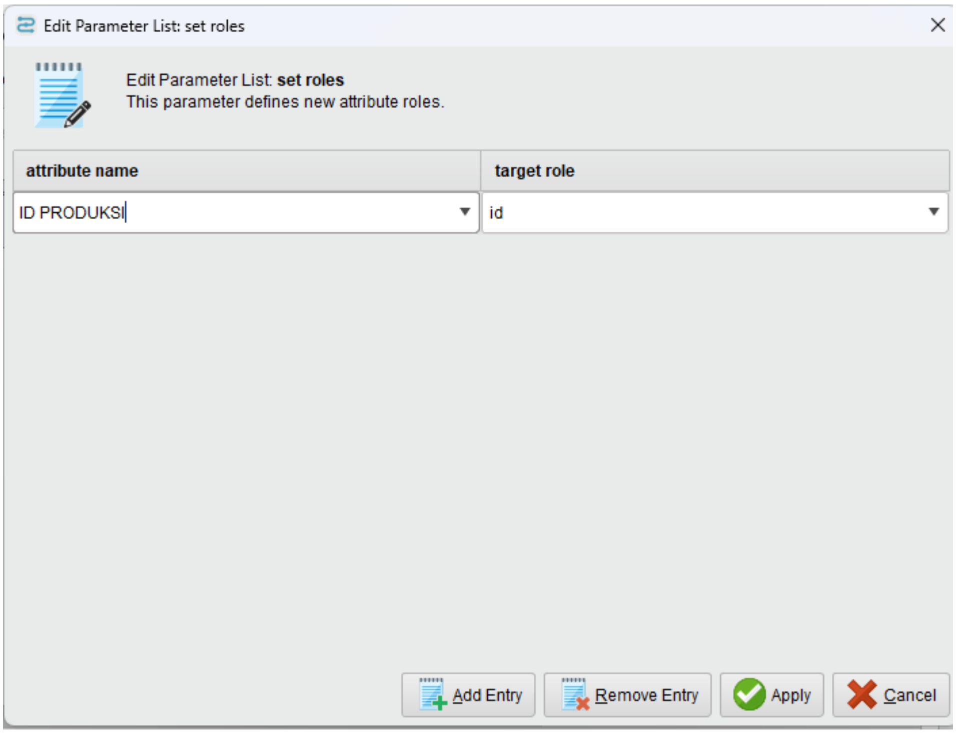


Figure 6: Proses Set Role

4.3. Transformation

Figure 7 below shows the use of the Nominal to Numeric Operator, which serves to convert categorical attributes in a production dataset, such as Product Type and Color, into a numerical format. This process is carried out by providing a unique number for each category of nominal attributes, so that the data can be processed by the K-Means algorithm which only supports numerical data analysis. This conversion ensures that each category is consistently represented in the analysis, making it easier for the algorithm to calculate the distance between the data in the clustering process. Using this operator, categorical attributes that were originally text are converted into numbers without losing their original meaning, allowing for more accurate grouping of data and according to the characteristics of the dataset.

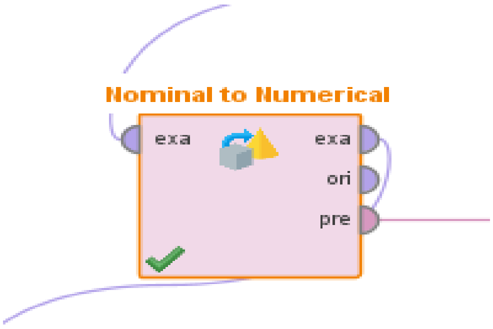


Figure 7: Operator Nominal To Numeric

In Figure 4.7 Nominal to Numerical Operator Parameters above, it works to convert categorical attributes, such as "Product Type" and "Color," into a numerical format by assigning a unique number to each category. This process is carried out so that nominal data can be converted into a numerical form that suits the needs of the K-Means algorithm, which can only work with numerical data. These conversions ensure that each category is consistently represented in the analysis, so that the algorithm can accurately process the data and produce optimal groupings. The results of the data transformation can be seen in figure 4.9 below.

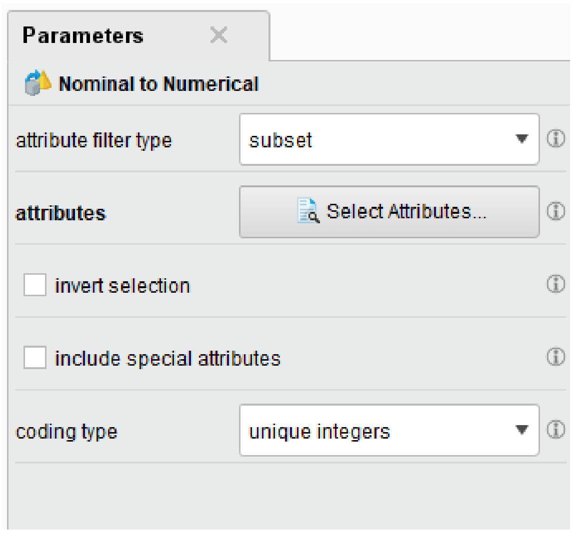


Figure 8: Parameter Operator Nominal

Figure 9 Numeric Representation The above column shows the results of converting the categorical attribute "Name" into numerical data through the Nominal to Numeric operator. At this stage, each unique value in the "Name" column is assigned a unique number as its numerical representation. For example, different product names, such as "Plain T-Shirts" and "Formal Shirts," are encoded into numbers like 1, 2, and so on. This process is carried out to ensure that nominal attributes can be understood and processed by the K-Means algorithm, which requires a numerical format to calculate the distance between the data in the clustering analysis./

Row No.	ID PRODUKSI	cluster ↑	NAMA PROD...	WARNA	UKURAN	BAHAN BAKU
98	C4RPEZ	cluster_1	3	16	2	4
99	C4RPEZ	cluster_1	3	16	3	4
100	C4RPEZ	cluster_1	3	16	4	4
101	C4RPEZ	cluster_1	3	17	0	4
102	C4RPEZ	cluster_1	3	17	1	4
103	C4RPEZ	cluster_1	3	17	2	4
104	C4RPEZ	cluster_1	3	17	3	4
105	C4RPEZ	cluster_1	3	17	4	4
106	C4RPEZ	cluster_1	3	18	0	4
107	C4RPEZ	cluster_1	3	18	1	4
108	C4RPEZ	cluster_1	3	18	2	4
109	C4RPEZ	cluster_1	3	18	3	4

ExampleSet (250 examples,2 special attributes,4 regular attributes)

Figure 9:. Numeric Representation of Columns

4.4. Data Mining

Figure 10 Below illustrates the use of the K-Means Algorithm operator to group convection production data at PT ABC based on attributes such as order quantity, product type, and raw materials. Cluster Distance Performance operators are used in this process to evaluate the quality of the clustering and determine the most optimal number of clusters (K). In this study, the K value is set at 3 to produce the production cluster that provides the best results.

Figure 10: K-Means

Figure 11 below shows the Cluster Distance Performance Operator in the K-Means algorithm which evaluates the distance between data in a cluster and measures how representative the cluster is to the analyzed data. This operator calculates the average distance between each data point in the cluster and its centroid, where a smaller distance indicates better cluster quality. Through this evaluation, operators help determine the optimal number of clusters by identifying differences in spacing between clusters, resulting in more structured and meaningful data clustering.

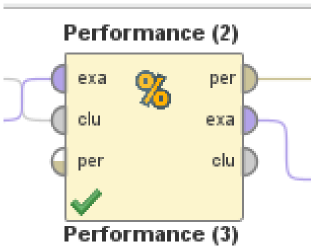


Figure 11: Operator Cluster Distance

Figure 12 shows the use of the Multiply operator in the data analysis process in this thesis. This operator serves to duplicate existing datasets or models, so that they can be used in several different analysis processes in parallel without changing the original data. In this study, the Multiply operator was used to create a copy of the dataset that had gone through the preprocessing and transformation stages, which was then applied to various other operators, such as the K-Means algorithm and cluster evaluation. This aims to ensure that each step of the analysis can be performed efficiently without affecting the main dataset, resulting in more organized and manageable results.

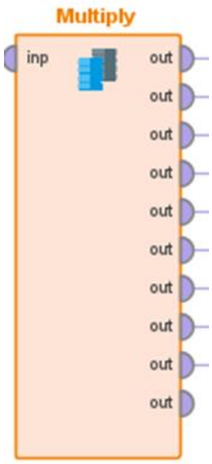


Figure 12: Operator Multiply

Figure 13 illustrates the evaluation process of each cluster generated by the K-Means algorithm. This evaluation was carried out to assess the quality of data grouping based on the distance between data in the cluster and the distance between different clusters. The evaluation method uses metrics such as the Silhouette Score or the Davies-Bouldin Index, which help identify how well the data in the cluster has internal similarities as well as how far the cluster differs from other clusters. This evaluation is important to ensure that the clusters generated are representative of patterns in production data, providing optimal clustering results and supporting strategic decision-making

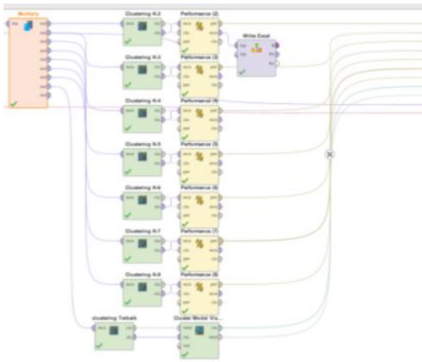


Figure 13: Evaluate Each Cluster

4.5. Evaluation & Interpretation

Table 3 below shows the Average Silhouette Width (ASW) value for each number of clusters generated from the clustering process using the K-Means algorithm. ASW values are used to evaluate the quality of a cluster by measuring how well the data in a cluster has internal similarities and how far the cluster differs from other clusters. The higher the ASW value, the better the quality of the resulting grouping. Based on the table, the cluster with the highest ASW value was chosen as the optimal number of clusters because it was able to produce the most representative grouping of production data patterns.

Table 3:

K	Davies Bouldin Index
2	0.539
3	0.655
4	0.765
5	0.928
6	0.880
7	0.852
8	0.836

Figure 14 shows the results of grouping production data using the K-Means algorithm for a model with the number of clusters K=2. In this model, the data is grouped into two clusters based on similarity of attributes, such as order quantity, product type, and raw materials. The clustering results illustrate a pattern that separates the data into two main groups: one group with high value attributes, and the other group with lower value attributes. This model provides a simple initial overview, but further analysis is needed to evaluate whether the number of clusters K=2 is optimal enough to represent the overall data pattern or if additional cluster counts are needed for more detailed results.

Cluster Model

```
Cluster 0: 125 items
Cluster 1: 125 items
Total number of items: 250
```

Figure 14: Cluster Model (K=2)

Figure 15 shows the results of the Performance Vector, which is used to evaluate the performance of the K-Means algorithm model in the clustering process. Performance Vector presents the value of evaluation metrics, such as the Davies-Bouldin Index and Silhouette Score, which are used to assess the quality of the grouping. The lower Davies-Bouldin Index value indicates that the clusters formed have a clear distance between the clusters and the data within the clusters have a high similarity. Meanwhile, a higher Silhouette Score indicates that the data in the cluster is more consistent and well separated from other clusters. Based on the results displayed, Performance Vector helps in determining the optimal number of clusters that provide the most representative grouping of production data and support strategic decision-making.

PerformanceVector

```
PerformanceVector:
Avg. within centroid distance: 31.208
Avg. within centroid distance_cluster_0: 29.735
Avg. within centroid distance_cluster_1: 32.682
Davies Bouldin: 0.539
```

Figure 15: Performance Vector

Figure 16 shows the results of grouping the data with the Best Cluster using the K-Means algorithm and the visualization generated by the Cluster Model Visualizer Operator. In this visualization, the best clusters are selected based on the evaluation of metrics such as the Silhouette Score or the Davies-Bouldin Index, which indicate the quality of the data grouping. The results show the division of production data into clusters that represent certain patterns, such as groups with high, medium, and low order volumes. This visualization also displays the centroid position of each cluster, which serves as an average representation of the attributes in that cluster. With the Operator Cluster

Model Visualizer, the analysis becomes easier to understand and helps in determining production strategies based on the characteristics of each cluster effectively.

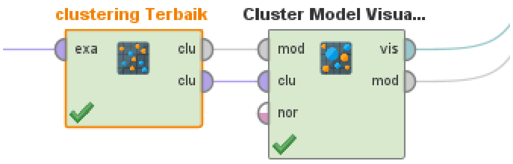


Figure 16: Best Cluster & Operator

Figure 17 shows the Cluster Visualization of the results of grouping production data using the K-Means algorithm. In this visualization, data that has been grouped into multiple clusters is displayed graphically to make it easier to interpret existing patterns. Each cluster is represented by a different color, making it easier to identify and analyze the distribution of data within each cluster. The centroid position is displayed as the center point of each cluster, which represents the average of the data attributes in that cluster. This visualization helps show how well the separation between clusters and the uniformity of the data within them are. By looking at these results, management can understand production patterns, such as product categories with high, medium, and low order volumes, which can be used to support strategic decision-making more efficiently.

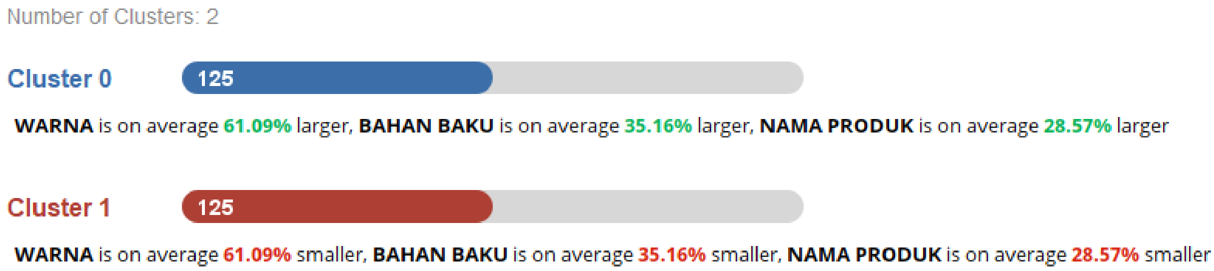


Figure 17: Cluster Visualization

Figure 18 shows the Heatmap Visualization, which is used to represent the pattern of grouping production data based on the K-Means algorithm. This heatmap depicts the intensity of the values of each attribute in the cluster, with a color indicating its level—the darker or lighter the color, the higher or lower the value of that attribute. This visualization helps to understand the distribution of data within a cluster, such as which attributes dominate or have the greatest influence in the formation of a particular cluster. For example, a cluster with a high order count might show more intense colors on related attributes, such as "Order" or "Product Type". With this heatmap, the analysis of data patterns becomes clearer, allowing management to identify production priorities and design a more focused strategy based on the characteristics of each cluster.

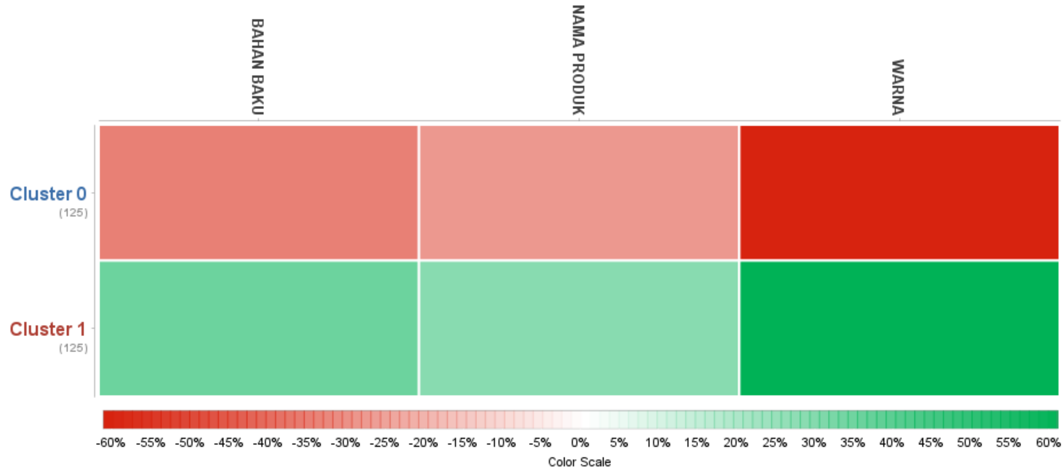


Figure 18: Heatmap Visualization

Figure 19 shows the Centroid of Each Cluster, which represents the mean point of the attributes in each cluster of clusters grouped using the K-Means algorithm. This centroid serves as the center of each cluster, describing the general characteristics of the data that is incorporated into that cluster. Each centroid shows an average value for attributes such as "Order Quantity," "Product Type," and "Raw Material," which helps in understanding the data patterns in each cluster. For example, a centroid with a high average order value can indicate a cluster of products with large demand. This centroid visualization makes it easier for management to identify focuses or priorities in production strategies, such as allocating more resources to high-order clusters or improving efficiency in low-order clusters. Thus, information from centroids becomes an important basis for data-driven decision-making.

Cluster	NAMA PRODUK	WARNA	UKURAN	BAHAN BAKU
Cluster 0	4.680	24.640	2.800	6.920
Cluster 1	2.600	5.952	2.400	3.320

Figure 19: Centroid Every Cluster

Figure 20 shows the Scatter Plot Visualization, which is used to visualize the results of the grouping of production data based on the K-Means algorithm. This scatter plot shows the distribution of data within each cluster, with each cluster represented by a different color to differentiate the data groups. In this scatter plot, the data points represent units of production data, such as the number of orders or other attributes, while the centroid position is displayed as a special sign indicating the center of each cluster. This visualization makes it easy to analyze data distribution patterns, such as seeing how tightly the data is in a cluster or how clear the separation between clusters is. This scatter plot is very useful for identifying data groups with similar characteristics, which can help management in designing production strategies, such as determining production priorities for high-order clusters or analyzing resource requirements for specific clusters. This visualization also provides intuitive insight into the effectiveness of the clustering process that has been carried out.

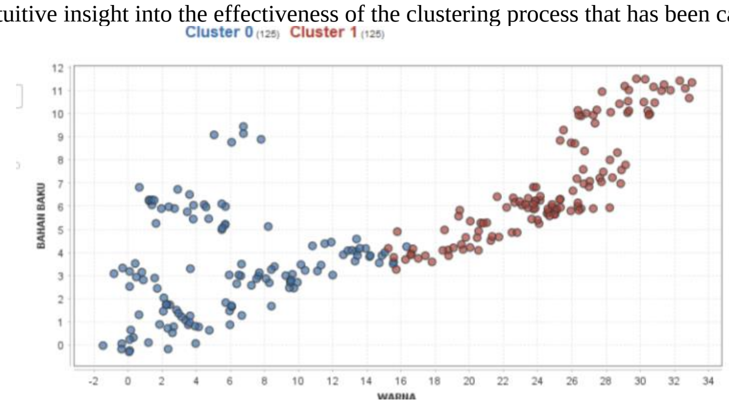


Figure 20: Scater Plot Visualization

A. Euclidean Distance Calculation
The formula of the Euclidean Distance is as follows:

$$d = \sqrt{\sum_{N} (X_i - Y_i)^2}$$

Known:
$X_1 = 3$
$X_2 = 16$
$X_3 = 2$
$X_4 = 4$

So:

Distance to Cluster 1 with Centroid = (2,600, 5,952, :2.400,3.320)

$$d = \sqrt{(3 - 2.600)^2 + (16 - 5.952)^2 + (2 - 2.400)^2 + (4 - 3.320)^2}$$

$$d = \sqrt{0.159 + 100.962 + 0.15 + 0.462}$$

$$d = \sqrt{101.733} = 10.086$$

Distance to Cluster 1 with Centroid = (, , , :4.68024.6402.8006.920)

$$d = \sqrt{(3 - 4.680)^2 + (16 - 24.640)^2 + (2 - 2.800)^2 + (4 - 6.920)^2}$$

$$d = \sqrt{2.822 + 74.629 + 0.639 + 8.526}$$

$$d = \sqrt{86.616} = 10.086$$

B. Data Clustering With Centroid

$$Centroid = \frac{Jumlah\ semua\ nilai\ attribut}{Jumlah\ data}$$

1. Cluster 0 (125 items):

a. Number of Students: 125

b. Total Number of Attribute Values:
   1) Total PRODUCT NAME Values = 325
   2) Total Value SUM =744
   3) Total PRICE Value = 300
   4) Total Values = 415

Centroid Cluster 0:

$$Centroid = \frac{325}{125}, \frac{744}{125}, \frac{300}{125}, \frac{415}{125}$$
$$= (2.6, 5.952, 2.4, 3.32)$$

2. Cluster 1 (1 items):

a. Number of Students: 125

b. Total Number of Attribute Values:
   1) Total Value PRODUCT NAME = 585
   2) Total Value SUM = 3080
   3) Total PRICE Value = 350
   4) Total Values = 865

Centroid Cluster 1:

$$Centroid = \frac{585}{125}, \frac{3080}{125}, \frac{350}{125}, \frac{865}{125}$$

$$= (4.68, 24.64, 2.8, 6.92)$$

5. Conclusion

This research has successfully implemented the K-Means Clustering algorithm to improve the production data clustering model in support of a more efficient convection production strategy. Based on the results of the analysis and evaluation, here are some conclusions that can be drawn:

1. The K-Means algorithm was successfully applied to PT ABC's production data using the Knowledge Discovery in Database (KDD) method, including the stages of selection, preprocessing, transformation, data mining, and evaluation. The data was successfully grouped into three clusters based on the similarity of attributes, such as the number of orders, product types, and raw materials. This process results in an effective clustering model, characterized by a low Davies-Bouldin Index evaluation value and a high Silhouette Score, which indicates good cluster quality.
2. The results of the grouping resulted in three main clusters, namely clusters with high, medium, and low order volumes. Analysis of patterns in clusters provides strategic insights for companies to prioritize resource allocation in high-order clusters and evaluate production efficiency in low-order clusters. The use of the K-Means algorithm has been proven to help companies in designing more efficient, responsive, and data-driven production strategies.

Acknowledgement

I would like to express my deepest gratitude to PT. ABC for the valuable permissions and support provided during the data collection process. Their cooperation is crucial to the success of this research. I am also grateful to my academic advisor for his insightful guidance and constructive input throughout this research. In addition, I really appreciate the contribution of all respondents who have provided the data needed for this study. Finally, I would like to express my sincere gratitude to my family and friends for their unrelenting support and enthusiasm, which greatly motivated me to complete this research.

Reference

[1] K. Aprilia and F. Sembiring, "Analysis of K-Means Algorithm on Spreading Report of the Cutting Department of Pt. Busana Indah Global," *JUSIM (Journal of Sist. Inf. Musirawas),* vol. 6, no. 2, pp. 159–168, 2021, doi: 10.32767/jusim.v6i2.1414.

[2] N. Ahsina, F. Fatimah, and F. Rachmawati, "Analysis of Bank Customer Segmentation Based on Credit Taking Using the K-Means Clustering Method," *J. Ilm. Technol. Applied Information.*, Vol. 8 No. 3 2022, doi: 10.33197/winner.Vol8.ISS3.2022.883.

[3] M. Azzam Al Fauzie and J. Akhir Putra, "Data Clustering Using the K-Means Method to Recommend Academic Learning for Active Students in Extracurriculars," *CLICK Review. Ilm. Inform. and Komput.*, vol. 4, no. 1, pp. 642–648, 2023, doi: 10.30865/click.v4i1.1116.

[4] R. Kurniawan, M. S. Hasibuan, and R. Hasibuan, "Clustering of Vaccine Priority Areas Using the K-Means Clustering Algorithm," *Media Online*, vol. 4, no. 3, pp. 1585–1592, 2023, doi: 10.30865/klik.v4i3.1334.

[5] Y. Filki, "K-Means Clustering Algorithm in Predicting Recipients of Direct Cash Assistance (BLT) for Village Funds," *J. Inform. Economist. Business*, vol. 4, pp. 166–171, 2022, doi: 10.37034/infeb.v4i4.166.

[6] A. R. Hayati, "Image Searching Data Color Image Clustering Algorithm K Means Clustering on Sales Data," *Skanika,* vol. 5, no. 2, pp. 203–215, 2022, doi: 10.36080/ska.v5i2.2943.

[7] F. P. A. Hasibuan, S. Sumarno, and I. Parlina, "Application of K-Means in Smartphone Product Sales Grouping," *SATESI J. Science and Technology. and Sist. Inf.*, vol. 1, no. 1, pp. 15–20, 2021, doi: 10.54259/satesi.v1i1.3.

[8] D. Haversyalapa, S. Puspasari, and R. Gustriansyah, "Pixel Image Clustering of Monpera Museum Photo Collection with the K-Means Method in Augmented Reality Applications," *IDEALIS Indones. J. Inf. Syst.*, vol. 7, no. 2, pp. 189–199, 2024, doi: 10.36080/idealis.v7i2.3175.

[9] K. F. G, Y. J. Sy, and P. Rahmawati, "Implementation of K-Means Algorithm on Clustering of Farmer Group Development for Increasing Corn Production," *J. Tek.*, vol. 15, no. 1, pp. 37–44, 2023, doi:10.30736/jt.v15i1.1012.

[10] Normah, B. Rifai, S. Vambudi, and R. Maulana, "Sentiment Analysis of Vtuber Development Using the SMOTE-Based Support Vector Machine Method," *J. Tech. Trembling. Amic BSI*, vol. 8, no. 2, pp. 174–180, 2022, doi:10.31294/jtk.v4i2.