Journal of Artificial Intelligence and Engineering Applications

*Website: https://ioinformatic.org/*

*15th June 2025. Vol. 4. No. 3; e-ISSN: 2808-4519*

# Application of K-Means for Clustering Analysis of Moring Sales in Stores Meowring

**Pramugya Jabhat Sakti[1*], Ade Irma Purnamasari[2], Agus Bahtiar[3], Edi Tohidi[4]**

*[1,2]Informatics Engineering, STMIK IKMI Cirebon, Indonesia*
*[3]Information System, STMIK IKMI Cirebon, Indonesia*
*[4]Computerized Accounting, STMIK IKMI Cirebon, Indonesia*
*pramugyajs@gmail.com[1*], irma2974@yahool.com[2], agusbahtir038@gmail.com[3] , editohidi00@ikmi.ac.id[4]*

**Abstract**

The complexity of stock management and marketing strategy at Toko Meowring requires a systematic analytical approach. This study implements the K-Means algorithm with a Knowledge Discovery in Databases (KDD) approach to optimize product segmentation. The analysis was conducted on 246 sales data over a year, considering product type, spiciness level, and sales volume. Evaluation using Davies-Bouldin Index (DBI) resulted in 7 optimal clusters with a DBI value of 0.402. The formed clusters identified bestseller product groups dominated by original flavors, low-demand products, moderate popularity products, stable sales products, and unique products with limited demand. This clustering enables optimization of inventory management, development of targeted promotional strategies, and improvement of product layout. The results validate the effectiveness of the K-Means algorithm in enhancing product segmentation accuracy for strategic decision-making.

***Keywords***: K-Means Clustering, Product Segmentation, Knowledge Discovery in Database, Davies-Bouldin Index, Inventory Management

## 1. Introduction

The rapid development of information technology has changed the way businesses operate, especially in terms of using data for decision-making [1]. One of the popular techniques in data analysis is data mining, specifically clustering algorithms that help businesses understand customer patterns and preferences. In the context of a retail business that sells unique products such as “moring” at Meowring stores, a deep understanding of product clustering and customer preferences is a challenge.

Although transaction data and customer reviews are available, many businesses still struggle to optimize the use of such data to increase sales and customer satisfaction. The k-means algorithm, as one of the clustering techniques, offers a solution to group products based on similar characteristics. However, the implementation of this algorithm faces various technical challenges such as the selection of relevant attributes and the determination of the optimal number of clusters [2].

Several previous studies have shown the successful implementation of k-means in retail sales data analysis. demonstrated the effectiveness of k-means in clustering products based on sales frequency and volume. used this technique to analyze sales of food and beverage products [3] while  successfully applied it for stock management and promotion strategies.

This research aims to improve the moring product sales clustering model using an optimized k-means algorithm based on sales data characteristics. The focus of the research is on analyzing numerical data using quantitative methods, with consideration of attributes such as purchase frequency and product category.The results of the study are expected to assist Meowring stores in developing more effective marketing strategies and contribute to the literature on the implementation of data mining in the context of local retail businesses in Indonesia[4].

Through this research, it is expected to gain a better understanding of the use of k-means algorithm in local product sales analysis, which can help similar businesses in optimizing their marketing strategies based on accurate and relevant data analysis [5] .

## 2. Literature Review

### 2.1. Data Mining

Testing by applying the K-means algorithm and Rapidminer tools to form a cluster model. Data mining is the process of analyzing data to find useful patterns, relationships, or hidden information from large data sets. This process is an important part of Knowledge Discovery in Database (KDD), which includes several stages, namely data selection, preprocessing, data transformation, the data mining process itself, and evaluation and interpretation of results [6].

In this research, data mining is used to process data on the education level of Bojong Village residents, which consists of attributes such as age, latest education, occupation, and marital status. Preprocessing stages are performed to clean the data, such as handling blank values and eliminating duplicates, so that the data is better prepared for analysis. Data transformation is then used to ensure the data has a uniform scale [7] .

### 2.2. Clustering

Clustering is a method in unsupervised learning that aims to group data into clusters based on the level of similarity between the data. This method is often used to understand the structure or distribution of data in a dataset without the need for labels or target variables [8]. Clustering enables the identification of patterns, such as the grouping of people based on education levels. This process helps in providing deeper insights to understand groups with similar characteristics, thus supporting more targeted decision making [9].

### 2.3. K-Means

K-Means is one of the most popular clustering algorithms. This algorithm works by dividing data into k groups based on the distance to the specified centroid. The K-Means process begins by selecting a value of k (the desired number of clusters), randomly assigning an initial centroid, then calculating the distance of the data to the centroid using Euclidean Distance. Data is clustered based on proximity to the centroid, and the centroid is updated until it reaches a steady state or maximum iteration [10].

## 3. Research Methods

This research uses the K-Means Clustering method to cluster sales data of moring products at Meowring Store. The research process is carried out through the Knowledge Discovery in Databases (KDD) approach which consists of the following stages:

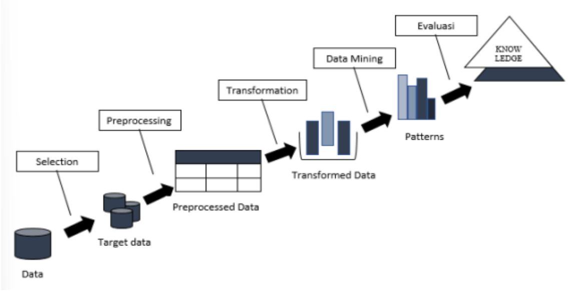


**Figure 1:** Research Methods

a. Data Selection;Sales data of moring products were collected from Meowring Store over a one-year period. The dataset used includes 246 data samples with 6 attributes, including product type and spiciness level (original, level 1 to level 5).
b. Preprocessing/Cleaning Data; The data is cleaned using the Replace Missing Values technique to remove blank or invalid values and ensure optimal data quality.
c. Transformation; Transformation is performed to prepare the data attributes so that they are suitable for the clustering process. The selected data includes product category attributes and sales amount.
d. Data Mining; The clustering process is performed with the help of RapidMiner software, where the k value (number of clusters) is tested using the Davies-Bouldin Index (DBI) and Elbow Method to determine the optimal cluster quality.
e. Data Interpretation (Evaluation); The quality of the clustering results was evaluated using DBI, where a low DBI value indicates better separation between clusters. The evaluation results show that 7 clusters are the optimal number with a DBI value of 0.402.

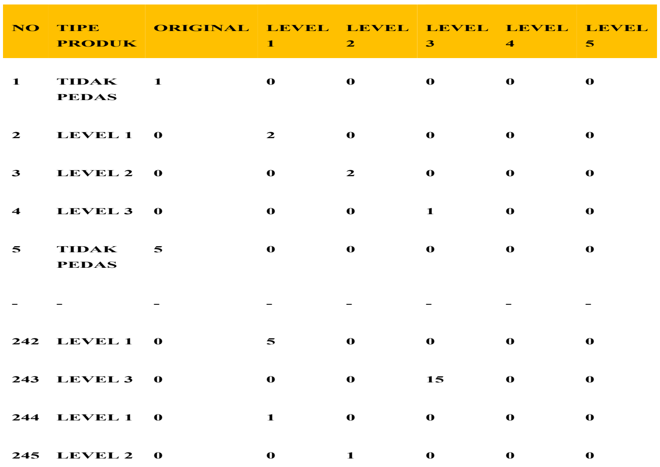
## 4. Results and Discussion

### 4.1. Research Result

#### 4.1.1. Data Selection

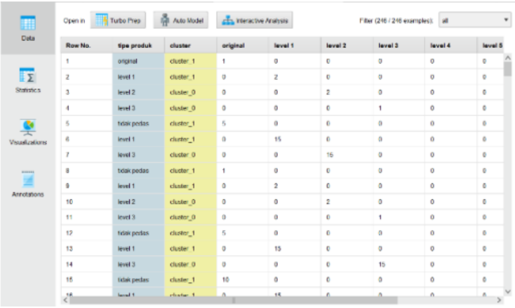
The initial stage in the Knowledge Discovery in Databases (KDD) process is Data Selection. The data used in this study is sales data for moring products at Meowring Stores collected over the past year. The dataset consists of 246 data samples with 6 main attributes, namely: NO PEDAS, LEVEL 1, LEVEL 2, LEVEL 3, LEVEL 4, and LEVEL 5.

The data selection and preparation process was carried out using Microsoft Excel to ensure the data was free from empty values, duplicates, or outliers. The dataset was then imported into RapidMiner using the Read Excel Operator so that the data could be processed further. Each attribute is categorized using the Set Role Operator to set the role of the attribute as an identity (ID) or analysis parameter.An example of a moring product sales dataset is presented in figure 2 below:



| NO | TIPE PRODUK | ORIGINAL | LEVEL 1 | LEVEL 2 | LEVEL 3 | LEVEL 4 | LEVEL 5 |
|---|---|---|---|---|---|---|---|
| 1 | TIDAK PEDAS | 1 | 0 | 0 | 0 | 0 | 0 |
| 2 | LEVEL 1 | 0 | 2 | 0 | 0 | 0 | 0 |
| 3 | LEVEL 2 | 0 | 0 | 2 | 0 | 0 | 0 |
| 4 | LEVEL 3 | 0 | 0 | 0 | 1 | 0 | 0 |
| 5 | TIDAK PEDAS | 5 | 0 | 0 | 0 | 0 | 0 |
| - | - | - | - | - | - | - | - |
| 242 | LEVEL 1 | 0 | 5 | 0 | 0 | 0 | 0 |
| 243 | LEVEL 3 | 0 | 0 | 0 | 15 | 0 | 0 |
| 244 | LEVEL 1 | 0 | 1 | 0 | 0 | 0 | 0 |
| 245 | LEVEL 2 | 0 | 0 | 1 | 0 | 0 | 0 |

**Figure 2:** Data Selection



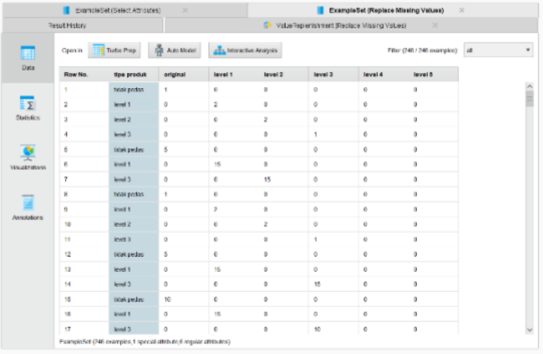
Filter (246 / 246 examples): all

| Row No. | tipe produk | cluster | original | level 1 | level 2 | level 3 | level 4 | level 5 |
|---|---|---|---|---|---|---|---|---|
| 1 | original | cluster_1 | 1 | 0 | 0 | 0 | 0 | 0 |
| 2 | level 1 | cluster_1 | 0 | 2 | 0 | 0 | 0 | 0 |
| 3 | level 2 | cluster_0 | 0 | 0 | 2 | 0 | 0 | 0 |
| 4 | level 3 | cluster_0 | 0 | 0 | 0 | 1 | 0 | 0 |
| 5 | tidak pedas | cluster_1 | 5 | 0 | 0 | 0 | 0 | 0 |
| 6 | level 1 | cluster_1 | 0 | 15 | 0 | 0 | 0 | 0 |
| 7 | level 3 | cluster_0 | 0 | 0 | 15 | 0 | 0 | 0 |
| 8 | tidak pedas | cluster_1 | 1 | 0 | 0 | 0 | 0 | 0 |
| 9 | level 1 | cluster_1 | 0 | 2 | 0 | 0 | 0 | 0 |
| 10 | level 2 | cluster_0 | 0 | 0 | 2 | 0 | 0 | 0 |
| 11 | level 3 | cluster_0 | 0 | 0 | 0 | 1 | 0 | 0 |
| 12 | tidak pedas | cluster_1 | 5 | 0 | 0 | 0 | 0 | 0 |
| 13 | level 1 | cluster_1 | 0 | 15 | 0 | 0 | 0 | 0 |
| 14 | level 3 | cluster_0 | 0 | 0 | 0 | 15 | 0 | 0 |
| 15 | tidak pedas | cluster_1 | 10 | 0 | 0 | 0 | 0 | 0 |
| 16 | level 1 | cluster_1 | 0 | 15 | 0 | 0 | 0 | 0 |

**Figure 3:** Set role

### 4.1.2. Preprocessing/ Cleaning

To ensure that the data is free of missing values so that it can be processed optimally in the clustering process. In this research, replacing missing values is done using the Replace Missing Value Operator in RapidMiner software.The results of this process show that all attributes of the dataset have been repaired, and there are no more empty values left. This ensures the dataset is ready to be used in the implementation of the K-Means Clustering algorithm.Figure 4 below displays the final result after the application of Replace Missing Values:



ExampleSet (Select Attributes)
ExampleSet (Replace Missing Values)
Result History
ValueReplenishment (Replace Missing Values)
Data
Statistics
Visualizations
Annotations
Open in
Turbo Prep
Auto Model
Interactive Analysis

Filter (246 / 246 examples): all

| Row No. | tipe produk | original | level 1 | level 2 | level 3 | level 4 | level 5 |
|---|---|---|---|---|---|---|---|
| 1 | tidak pedas | 1 | 0 | 0 | 0 | 0 | 0 |
| 2 | level 1 | 0 | 2 | 0 | 0 | 0 | 0 |
| 3 | level 2 | 0 | 0 | 2 | 0 | 0 | 0 |
| 4 | level 3 | 0 | 0 | 0 | 1 | 0 | 0 |
| 5 | tidak pedas | 5 | 0 | 0 | 0 | 0 | 0 |
| 6 | level 1 | 0 | 15 | 0 | 0 | 0 | 0 |
| 7 | level 3 | 0 | 0 | 15 | 0 | 0 | 0 |
| 8 | tidak pedas | 1 | 0 | 0 | 0 | 0 | 0 |
| 9 | level 1 | 0 | 2 | 0 | 0 | 0 | 0 |
| 10 | level 2 | 0 | 0 | 2 | 0 | 0 | 0 |
| 11 | level 3 | 0 | 0 | 0 | 1 | 0 | 0 |
| 12 | tidak pedas | 5 | 0 | 0 | 0 | 0 | 0 |
| 13 | level 1 | 0 | 15 | 0 | 0 | 0 | 0 |
| 14 | level 3 | 0 | 0 | 0 | 15 | 0 | 0 |
| 15 | tidak pedas | 10 | 0 | 0 | 0 | 0 | 0 |
| 16 | level 1 | 0 | 15 | 0 | 0 | 0 | 0 |
| 17 | level 3 | 0 | 0 | 0 | 10 | 0 | 0 |

ExampleSet (246 examples,1 special attribute,6 regular attributes)

**Figure 4:** Data processing results

### 4.1.3. Transfomation

The Data Transformation stage aims to prepare the dataset to be processed using the K-Means Clustering algorithm. The transformation process is carried out by changing the data structure into a format suitable for analysis.

### 4.1.4. Data Mining

Sales data of moring products that have been cleaned are processed using the K-Means Clustering Operator in RapidMiner. The initial parameter k (number of clusters) is set starting from k = 2 until a maximum of 10 iterations are performed. This process ensures that the data can be optimally grouped according to the pattern formed. This stage is shown in Figure below:

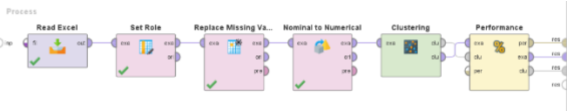


**Figure 5:** K-Means Clustering

After the K-Means process is applied, the performance of the model is evaluated with the Performance Operator. This evaluation involves measuring the distance between clusters and the Davies-Bouldin Index (DBI) value to assess the quality of separation between clusters. This stage can be seen in Figure 7.

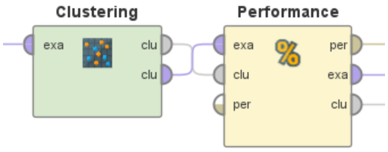
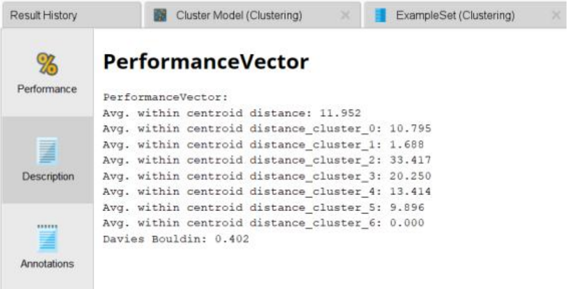


**Figure 6:** Prefomance

The evaluation results show the average value of the distance between centroids (Avg) which reflects the efficiency and effectiveness of data transformation. This value is used to determine the optimal clustering quality. The average results can be seen in Figure 8.



PerformanceVector

```
PerformanceVector:
Avg. within centroid distance: 11.952
Avg. within centroid distance_cluster_0: 10.795
Avg. within centroid distance_cluster_1: 1.688
Avg. within centroid distance_cluster_2: 33.417
Avg. within centroid distance_cluster_3: 20.250
Avg. within centroid distance_cluster_4: 13.414
Avg. within centroid distance_cluster_5: 9.896
Avg. within centroid distance_cluster_6: 0.000
Davies Bouldin: 0.402
```

**Figure 7:** Evaluation Avg Centroid

### 4.1.5. Interpretation / Evaluation

The Model Evaluation stage is conducted to assess the quality of the clustering results using the Davies-Bouldin Index (DBI). The DBI value is used as the main indicator to determine the optimal number of clusters, where a lower DBI value indicates better cluster separation. The evaluation process was carried out by trying to vary the number of clusters k from k=2 to k=10. The evaluation results can be seen in Table 1 below:

**Table 2:** Evaluation DBI Results

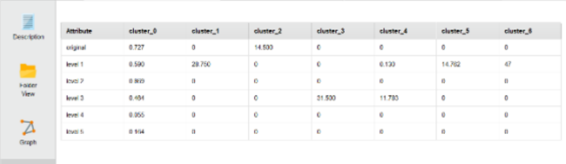
| k | DBI | Cluster |
|---|---|---|
| 2 | Davies<br>Bouldin:<br>0.598 | Cluster 0: 220 items<br>Cluster 1: 26 items<br>Total number of items: 246 |
| 3 | Davies<br>Bouldin:<br>0.562 | Cluster 0: 202items<br>Cluster 1: 22 items<br>Cluster 2: 22 items<br>Total number of items: 246 |
| 4 | Davies<br>Bouldin:<br>0.552 | Cluster 0: 190 items<br>Cluster 1: 22 items<br>Cluster 2: 12 items<br>Cluster 3: 22 items<br>Total number of items: 246 |
| 5 | Davies<br>Bouldin:<br>0.526 | Cluster 0: 186 items<br>Cluster 1: 5 items<br>Cluster 2: 22 items<br>Cluster 3: 21 items<br>Cluster 4: 12 items<br>Total number of items: 246 |
| 6 | Davies<br>Bouldin:<br>0.501 | Cluster 0: 5 items<br>Cluster 1: 183 items<br>Cluster 2: 2 items<br>Cluster 3: 23 items<br>Cluster 4: 21 items<br>Cluster 5: 12 items<br>Total number of items: 246 |
| 7 | Davies<br>Bouldin:<br>0.402 | Cluster 0: 183 items<br>Cluster 1: 4 items<br>Cluster 2: 12 items<br>Cluster 3: 2 items<br>Cluster 4: 23 items |

| | | |
|---|---|---|
| | | Cluster 5: 21 items<br>Cluster 6: 1 items<br>Total number of items: 246 |
| 8 | Davies Bouldin: 0.510 | Cluster 0: 17 items<br>Cluster 1: 24 items<br>Cluster 2: 5 items<br>Cluster 3: 2 items<br>Cluster 4: 150 items<br>Cluster 5: 21 items<br>Cluster 6: 4 items<br>Cluster 7: 23 items<br>Total number of items: 246 |
| 9 | Davies Bouldin: 0.500 | Cluster 0: 17 items<br>Cluster 1: 24 items<br>Cluster 2: 7 items<br>Cluster 3: 2 items<br>Cluster 4: 150 items<br>Cluster 5: 18 items<br>Cluster 6: 4 items<br>Cluster 7: 23 items<br>Cluster 8: 1 items<br>Total number of items: 246 |
| 10 | Davies Bouldin: 0.422 | Cluster 0: 145 items<br>Cluster 1: 2 items<br>Cluster 2: 4 items<br>Cluster 3: 4 items<br>Cluster 4: 17 items<br>Cluster 5: 21 items<br>Cluster 6: 1 items<br>Cluster 7: 24 items<br>Cluster 8: 10 items<br>Cluster 9: 18 items<br>Total number of items: 246 |

### 4.1.6. Knowledge

#### a) Evaluation Model

After determining k=7 as the optimal number of clusters, the clustering results were interpreted based on the centroid value for each attribute (original, level 1, level 2, level 3, level 4, and level 5). The centroid results show the contribution of each attribute to the clusters formed. The centroid value results are shown in Figure 9 below:



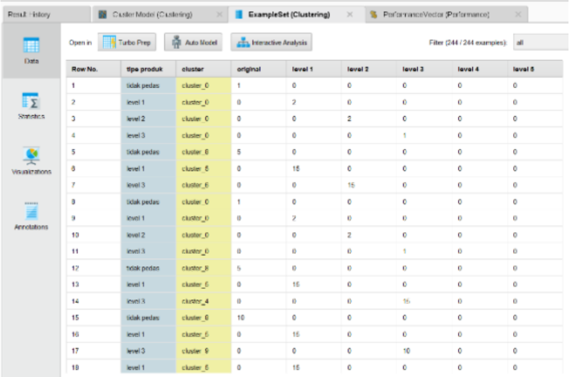
| Attribute | cluster_0 | cluster_1 | cluster_2 | cluster_3 | cluster_4 | cluster_5 | cluster_6 |
|---|---|---|---|---|---|---|---|
| original | 0.727 | 0 | 14.500 | 0 | 0 | 0 | 0 |
| level 1 | 0.590 | 28.750 | 0 | 0 | 0.133 | 14.762 | 47 |
| level 2 | 0.869 | 0 | 0 | 0 | 0 | 0 | 0 |
| level 3 | 0.454 | 0 | 0 | 31.500 | 11.783 | 0 | 0 |
| level 4 | 0.055 | 0 | 0 | 0 | 0 | 0 | 0 |
| level 5 | 0.164 | 0 | 0 | 0 | 0 | 0 | 0 |

**Figure 8:** Evaluation Table Centroid

Based on the centroid value, the interpretation of the clustering results is as follows:

- Cluster_0 has the highest contribution to the original attribute with a value of 0.727.
- Cluster_1 shows significant dominance on level 1 attributes with a value of 28.750, followed by Cluster_5 with a value of 14.762.
- Cluster_3 dominates level 3 attributes with a value of 31,500, followed by Cluster_4 with a value of 11,783.

Visualization of the results of grouping data into 7 clusters is shown in Figure 10.



Filter (244 / 244 examples): all

| Row No. | tipe produk | cluster | original | level 1 | level 2 | level 3 | level 4 | level 5 |
|---|---|---|---|---|---|---|---|---|
| 1 | tidak pedas | cluster_0 | 1 | 0 | 0 | 0 | 0 | 0 |
| 2 | level 1 | cluster_0 | 0 | 2 | 0 | 0 | 0 | 0 |
| 3 | level 2 | cluster_0 | 0 | 0 | 2 | 0 | 0 | 0 |
| 4 | level 3 | cluster_0 | 0 | 0 | 0 | 1 | 0 | 0 |
| 5 | tidak pedas | cluster_6 | 5 | 0 | 0 | 0 | 0 | 0 |
| 6 | level 1 | cluster_6 | 0 | 15 | 0 | 0 | 0 | 0 |
| 7 | level 3 | cluster_6 | 0 | 0 | 15 | 0 | 0 | 0 |
| 8 | tidak pedas | cluster_0 | 1 | 0 | 0 | 0 | 0 | 0 |
| 9 | level 1 | cluster_0 | 0 | 2 | 0 | 0 | 0 | 0 |
| 10 | level 2 | cluster_0 | 0 | 0 | 2 | 0 | 0 | 0 |
| 11 | level 3 | cluster_0 | 0 | 0 | 0 | 1 | 0 | 0 |
| 12 | tidak pedas | cluster_8 | 5 | 0 | 0 | 0 | 0 | 0 |
| 13 | level 1 | cluster_6 | 0 | 15 | 0 | 0 | 0 | 0 |
| 14 | level 3 | cluster_4 | 0 | 0 | 0 | 15 | 0 | 0 |
| 15 | tidak pedas | cluster_6 | 10 | 0 | 0 | 0 | 0 | 0 |
| 16 | level 1 | cluster_6 | 0 | 15 | 0 | 0 | 0 | 0 |
| 17 | level 3 | cluster_9 | 0 | 0 | 0 | 10 | 0 | 0 |
| 18 | level 1 | cluster_6 | 0 | 15 | 0 | 0 | 0 | 0 |

**Figure 9:** Evaluation ExampleSet Clustering

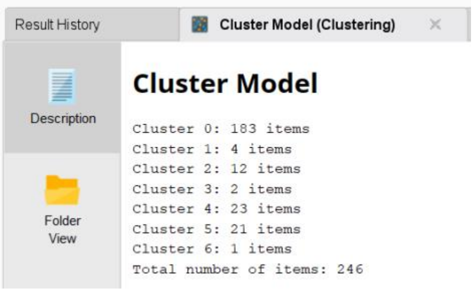
With these results, the sales data of moring products were successfully grouped into 7 clusters based on sales patterns and product spiciness levels. This interpretation provides insight into the characteristics of each product group that can be used to support decision making, such as marketing strategies and stock management.

**b) Evaluation**

This research uses the K-Means Clustering algorithm with RapidMiner tools to cluster moring product sales data. Model evaluation was conducted using two approaches, namely Davies-Bouldin Index (DBI) and Elbow Method. The complete evaluation results can be seen as follows:

1. Davies-Bouldin Index (DBI) Value; Model evaluation is performed with the Davies-Bouldin Index (DBI) to determine the optimal number of clusters. DBI evaluates the extent to which clusters are optimally separated. From the test results with variations in the number of clusters k=2 to k=10, the smallest DBI value is obtained at k=7, which indicates the best cluster quality.
   At k=7, the following results were obtained:
   - DBI value: 0.423
   - Cluster Member Distribution:
   - Cluster_0: 183 items
   - Cluster_1: 4 items
   - Cluster_2: 12 items
   - Cluster_3: 2 items
   - Cluster_4: 23 items
   - Cluster_5: 21 items
   - Cluster_6: 1 item

These results show that clustering with 7 clusters results in optimal separation between clusters. Visualization of the clustering model is shown in Figure 11 below:



Cluster Model

```
Cluster 0: 183 items
Cluster 1: 4 items
Cluster 2: 12 items
Cluster 3: 2 items
Cluster 4: 23 items
Cluster 5: 21 items
Cluster 6: 1 items
Total number of items: 246
```

**Figure 10:** Evaluation Cluster Model

1. Elbow Method; The Elbow method is used to validate the optimal number of clusters by looking at changes in the value of Avg. Within Centroid Distance. This value describes the average distance between points in the cluster with respect to its centroid. The Elbow graph is obtained by plotting the average within-cluster distance for each value of k. The elbow point on the graph indicates the optimal number of clusters. In this study, the elbow point occurs at k=7, which indicates that 7 clusters is the optimal number. The similarity of results between the DBI evaluation and the Elbow method reinforces this conclusion.

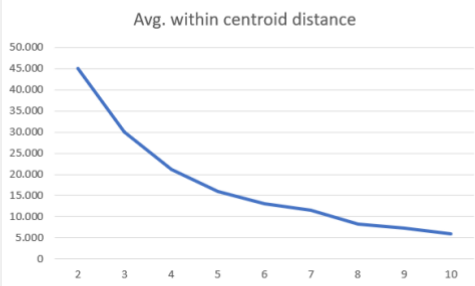


**Figure 11:** Method Elbow

2. Center Distribution Plot; The clustering result plot illustrates the distribution of product types based on spiciness level (original to level 5) and number of sales. This plot is useful for providing an in-depth visualization of the dominance and distribution of product categories. The results of the cluster distribution plot are shown in Figure 13 below:

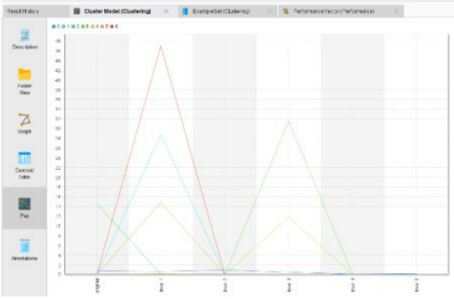
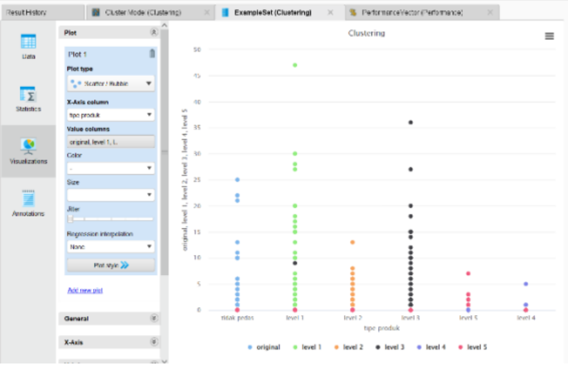


**Figure 12:** Result Plot

The analysis results show:
- Level 2 (red color) has the highest peak, indicating the dominance of products at that level.
- Level 3 (green color) also displays significant contribution with the highest peak in this category.
- The original category and other levels 1-5 have smaller contributions and are spread across several levels.

3. Clustering Data Visualization; The results of data visualization using scatter plots provide an overview of data distribution based on product type (X-axis) and numerical sales value (Y-axis). Each point on the scatter plot represents sales data and is given a different color to distinguish product categories.



Plot type: Scatter / Bubble
X-Axis column: tipe produk
Value columns: original, level 1, l...
Color: -
Size:
Jitter
Regression interpolation: None

**Figure 13:** Clustering Visualization Results

This visualization shows:
- The “not spicy” category only has data in the original with a small spread of values.
- In level 2 and level 3 categories, the distribution of values shows more significant dominance than other categories.
- The sequences at level 4 and level 5 show a distribution of data that needs further study for clarification.

### 4.2. Discussion

a) Moring Product Sales Clustering Model Analysis Results Based on the application of the K-Means Clustering
Based on the application of the K-Means Clustering algorithm using RapidMiner, evaluation is performed with the Davies-Bouldin Index (DBI) and variations in the value of k from 2 to 10. The evaluation results show that the smallest DBI value is obtained at k = 7 with a DBI value of 0.402. This value indicates that the separation between clusters at k=7 has optimal quality.
The distribution of the number of items at k=7 is as follows:
- Cluster 0: 183 items (high sales)
- Cluster 1: 4 items (very low sales)
- Cluster 2: 12 items (medium sales)
- Cluster 3: 2 items (very low sales)
- Cluster 4: 23 items (stable sales)
- Cluster 5: 21 items (stable sales)
- Cluster 6: 1 item (unique or specific product).

These clustering results show that there are variations in product sales that can be grouped into 7 clusters to provide a more in-depth picture of sales behavior at Meowring stores.

b) Clustering Results for Marketing Strategy.
The results of the clustering analysis provide an in-depth picture of the characteristics of product sales performance at Meowring Store. The interpretation of each cluster is as follows:
1) Cluster 0: Products with the highest sales, dominated by original variants and low spiciness. These products represent customers' top preferences and should be prioritized for availability.
2) Cluster 1: Products with very low sales. The strategy required is intensive promotion such as discounts or product bundling to increase appeal.
3) Cluster 2: Products with medium sales that have the potential to be increased through seasonal promotions or special marketing campaigns.
4) Cluster 3: Products with the least sales. Need to re-evaluate regarding quality, relevance, and market demand.
5) Cluster 4: Products with stable sales, showing consistent performance. Customer loyalty strategies can help maintain this stability.
6) Cluster 5: Products with stable sales similar to the previous cluster, suitable for maintaining adequate stock to meet the needs of loyal customers.
7) Cluster 6: Unique or specific products with very small sales contribution. In-depth analysis is required to understand the relevance and market opportunities of these products

## 5. Conclusion

Based on the results of research and analysis that has been carried out on sales data of moring products at Meowring stores using the K-Means Clustering algorithm with RapidMiner tools, it is found that the optimal number of clusters is 7 clusters with the smallest Davies-Bouldin Index (DBI) value of 0.402. This evaluation is reinforced by the application of the Elbow Method which shows an elbow point at k=7. Of the seven clusters formed, Cluster 0 shows the highest selling products dominated by original variants and low spiciness levels, while Clusters 1 and 3 show products with very low sales that require special promotional strategies. Cluster 2 includes products with medium sales that have the potential to be improved, while Clusters 4 and 5 show products with stable sales. Cluster 6 identifies unique or specific products with very small sales contributions that require further analysis.

Based on these clustering results, some marketing strategies that can be implemented include stock prioritization for popular products, increasing product attractiveness through promotions or discounts for low-selling products, identifying seasonal promotional opportunities for potential products, as well as maintaining stable product availability to support customer loyalty. With the application of this K-Means Clustering algorithm, Meowring stores can optimize stock management, design more targeted marketing strategies, and improve customer loyalty.

## References

[1] Abdalla, H. B. (2022). A brief survey on big data: technologies, terminologies and data-intensive applications. *Journal of Big Data*, *9*(1), 1–36. https://doi.org/10.1186/S40537-022-00659-3/TABLES/5

[2] Awaliah, L., Rahaningsih, N., & Dana, R. D. (2024). Implementasi Algoritma K-Means Dalam Analisis Cluster Korban Kekerasan Di Provinsi Jawa Barat. *JATI (Jurnal Mahasiswa Teknik Informatika)*, *8*(1), 188–195. https://doi.org/10.36040/jati.v8i1.8332

[3] Dilawati, H., Widianto, H., & Kuswiadji, A. (2024). *Klasterisasi Data Rekam Medis Pasien Menggunakan Metode K-Means Clustering Di Rumah Sakit Widodo Ngawi*. *5*(2), 139–147.

[4] Ezugwu, A. E., Shukla, A. K., Agbaje, M. B., Oyelade, O. N., José-García, A., & Agushaka, J. O. (2021). Automatic clustering algorithms: a systematic review and bibliometric analysis of relevant literature. *Neural Computing and Applications*, *33*(11), 6247–6306. https://doi.org/10.1007/S00521-020-05395-4/METRICS

[5] Febrianty, E., Awalina, L., & Rahayu, W. I. (2023). Optimalisasi Strategi Pemasaran dengan Segmentasi Pelanggan Menggunakan Penerapan K-Means Clustering pada Transaksi Online Retail. *Jurnal Teknologi Dan Informasi*, *13*(2), 122–137. https://doi.org/10.34010/JATI.V13I2.10090

[6] Fitri, E. M., Suryono, R. R., & Wantoro, A. (2023). KLASTERISASI DATA PENJUALAN BERDASARKAN WILAYAH MENGGUNAKAN METODE K-MEANS PADA PT XYZ. *Jurnal Komputasi*, *11*(2), 157–168. https://doi.org/10.23960/KOMPUTASI.V11I2.12582

[7] Gupta, M. K., & Chandra, P. (2020). A comprehensive survey of data mining. *International Journal of Information Technology (Singapore)*, *12*(4), 1243–1257. https://doi.org/10.1007/S41870-020-00427-7/METRICS

[8] Gutierrez-Osorio, C., & Pedraza, C. (2020). Modern data sources and techniques for analysis and forecast of road accidents: A review. *Journal of Traffic and Transportation Engineering (English Edition)*, *7*(4), 432–446. https://doi.org/10.1016/J.JTTE.2020.05.002

[9] Hartati, T., Nurdiawan, O., Wiyandi, E., & Cirebon, S. I. (2021). Analisis dan Penerapan Algoritma K-Means Dalam Strategi Promosi Kampus Akademi Maritim Suaka Bahari. *Jurnal Sains Teknologi Transportasi Maritim*, *3*(1), 1–7. https://doi.org/10.51578/J.SITEKTRANSMAR.V3I1.31

[10] Ikotun, A. M., Almutari, M. S., & Ezugwu, A. E. (2021). K-Means-Based Nature-Inspired Metaheuristic Algorithms for Automatic Data Clustering Problems: Recent Advances and Future Directions. *Applied Sciences 2021, Vol. 11, Page 11246*, *11*(23), 11246. https://doi.org/10.3390/APP112311246