Bulletin of Network Engineer and Informatics
E-ISSN 2986-8017 | P-ISSN 2987-4858
https://bufnets.tech

Vol. 4 No 1 – April 2026
DOI: 10.59688/fqnk3355

# OPTIMIZATION GROUPING QUALITY MBG PROGRAM FOOD USES K-MEANS ALGORITHM BASED ON DAVIES-BOULDIN INDEX EVALUATION

**Fitriasih[1], Fahrudin[2], Heny Indriani[3],Denny vasanandosabanise[4], Muhammad Ainurrohman[5]**

Purbaya Polytechnic
Tegal, Indonesia

e-mail: fitriasih@purbaya.ac.id[1], fahrudinmkom@gmail.com[2], heny indriani@purbaya.ac.id[3], denny.vasanando@gmail.com[4], muhammadainurrohman@purbaya.ac.id[5]

| Received : 10 April 2026 | Accepted : 29 April 2026 | Published : 03 May 2026 |
|---|---|---|

***Abstract***

*The evaluation of food quality in the Free Nutritious Meal Program (MBG) requires an objective and systematic approach to ensure nutritional standards and service effectiveness. This study applies a data mining technique using the K-Means clustering algorithm to classify food quality based on multiple attributes, including Calories, Protein, Fat, Carbohydrates, Freshness, Cleanliness, Serving Temperature, and Eligibility. This research utilizes RapidMiner to perform data preprocessing, involving data preprocessing through normalization, clustering with K-Means, and performance evaluation using the Davies-Bouldin Index (DBI) and Average Within Centroid Distance. Two clustering scenarios, K=3 and K=5, were evaluated to determine the optimal number of clusters.The results indicate that the K=3 model achieves a lower DBI value (0.214) compared to K=5 (0.224), indicating better cluster separation. Although K=5 produces more compact clusters, it does not improve overall clustering quality due to weaker separation. Therefore, the K=3 configuration is identified as the optimal model, as it provides a better balance between cluster separation and interpretability. These findings demonstrate that a multi-attribute clustering approach can effectively support data-driven decision-making in evaluating and improving food quality in the MBG program.*

***Keywords:*** *K-Means Clustering, Food Quality, Data Mining, Davies-Bouldin Index, Multi-Attribute Analysis.*

## 1. INTRODUCTION

Data mining is the process of extracting patterns and information hidden from a set of data in large amounts using statistical techniques, machine learning, and database management [1]. One of the methods in data mining is clustering, namely the technique of grouping data based on the level of similarity characteristics between objects without using class labels previously [2]. In the context of the Free Nutritional Meal Program (MBG), the quality of food served needs to be evaluated in a way periodically to ensure food quality standards are still maintained. With the increasing amount of food evaluation data, an automatic method is necessary to group food quality objectively and systematically.

K-Means algorithm is one of the most widely used clustering methods used because of its own ability fast and effective computing in grouping numeric data based on centroid distance [3]. Some previous studies show that effective K-Means algorithm used For quality data grouping product and food [4] –[ 6]. However, one of the main challenges in the use of K-Means is determining the optimal number of clusters. Therefore, it is necessary to evaluate using the Davies-Bouldin Index (DBI) to evaluate quality cluster separation and Average Within Centroid Distance to measure cluster cohesiveness [7].

## 2. RESEARCH METHODOLOGY

This study uses Knowledge Discovery in Database (KDD) approach which includes stages data selection, preprocessing, transformation, data mining, and evaluation [1]. The research dataset consists of 500 samples of food quality MBG Program food collected from results observation assessment of food menus in the MBG program.

### 2.1 Dataset Description
The dataset consists of the following attributes:

Table 1: Dataset Details
[Source, Author, 2026]

| No | Attribute | Description |
|---|---|---|
| 1 | ID_Menu | Unique code for every food menu |
| 2 | Menu_Name | food menu |
| 3 | Calories | Total food energy (kcal ) |
| 4 | Protein_g | Protein content in grams |
| 5 | Fat_g | Fat content in grams |
| 6 | Carbohydrate_g | Contents carbohydrates in grams |
| 7 | Freshness_1_10 | Food freshness score (1–10) |
| 8 | Cleanliness_1_10 | Cleanliness score food (1–10) |
| 9 | Serving Temperature_C | Temperature presentation of food in Celsius |
| 10 | Eligibility_1_10 | Eligibility score consumption food (1–10) |

### 2.2 Data Preprocessing
preprocessing stage is carried out To ensure data is ready to be processed in clustering algorithm. Stages This includes data cleaning, normalization, and selection attributes.

#### 2.2.1 Data Cleaning
At the stage This is done inspection against duplicate data, empty data (missing values), as well as inconsistency in data writing format. Data that is incomplete or invalid fixed and deleted so as not to influence clustering results.

#### 2.2.2 Normalization Using Z-Transformation
After the data cleaning process, it is carried out normalization using Z-Transformation (Z-Score Normalization) method. This method is used to standardize all over attribute numeric so that it has a mean of 0 and a standard deviation of 1, so that each attribute has its own balanced contribution in the calculation process Euclidean distance [8].

Z-Transformation normalization formula: Z = (X - mean) / standard deviation
Information:

- X = actual data value
- Mean = average of attributes
- Standard Deviation = standard deviation attribute

The use of Z-Transformation is chosen because the dataset has a range of mark different attributes, such as mark calories, temperature presentation and scores assessment, ensuring that attributes with larger numerical scales do not dominate the clustering process.

#### 2.2.3 Attribute Selection
At the attribute selection stage, this study utilizes all relevant attributes in the dataset for the clustering analysis process. The attributes used include Calories, Protein_g, Fat_g, Carbohydrate_g, Freshness_1_10, Cleanliness_1_10, Serving_Temperature_C, and Eligibility_1_10, which collectively represent the nutritional content, hygiene, and overall quality of MBG program food.

The use of multiple attributes is intended to provide a comprehensive and multidimensional representation of food quality, as each variable contributes important information. Nutritional attributes such as Calories, Protein, Fat, and Carbohydrates reflect the energy and nutritional value of the food, while Freshness, Cleanliness, Serving Temperature, and Eligibility represent qualitative aspects related to food safety and suitability for consumption.

Unlike approaches that rely on a single attribute, this study adopts a multi-attribute approach to improve the accuracy, robustness, and reliability of the clustering results. By incorporating all relevant variables, the clustering process is able to capture more complex patterns and relationships within the data.
In this research, attribute selection is not performed to eliminate variables, but rather to ensure that all important attributes are included in the analysis. The normalization process using Z-Transformation ensures that each attribute contributes proportionally and prevents attributes with larger numerical ranges from dominating the clustering process.

The implementation in RapidMiner is carried out without restricting specific attributes using the *Select Attributes* operator, allowing the K-Means algorithm to process all variables simultaneously. This approach is expected to produce more representative clustering results and better reflect the real characteristics of food quality in the MBG program.
.

**2.3 Clustering Process**

The clustering method used is K-Means algorithm with two scenarios number of clusters, namely K=3 and K=5. Algorithm Work in an iterative way through stages following :

1. Determine number of clusters (K).
2. Initialize the initial centroid in a random way.
3. Calculate distance of each data to the centroid using Euclidean Distance.
4. Grouping data to the nearest centroid.
5. Count repeat centroid positions based on the average of cluster members.
6. Repeat the process until the centroid is stable. or No occur cluster changes.

**2.4 Cluster Evaluation**

Stage cluster evaluation is carried out For evaluating quality grouping results produced by the K-Means algorithm and For determining optimal number of clusters in a quality dataset MBG Program food. Evaluation this is very important because the clustering process is Unsupervised learning methods do not have class labels reference, so that required metric internal validation to measure how much Good the cluster structure formed [7]. In this research, evaluation of cluster quality is performed using two metrics internal validation, namely Davies-Bouldin Index (DBI) and Average Within Centroid Distance.

**2.4.1 Davies-Bouldin Index (DBI)**

The Davies-Bouldin Index is used to measure quality separation between clusters based on comparison between intra-cluster and inter-cluster distances. Metrics This calculates the average similarity between each cluster with the most similar other cluster, where the similarity is determined based on the ratio distance between centroid and dispersion cluster members.
A smaller DBI value indicates superior clustering quality, reflecting higher separation and minimal overlap. Small DBI value indicates that each cluster has more characteristics clear and distinct from each other. In the context of this study, DBI is used as the main indicator to determine the best number of clusters. Because the focus of the study is getting grouping quality food that has a clear separation limit categories between high quality, medium, and low.

**2.4.2 Average Within Centroid Distance**

Average Within Centroid Distance is used to measure the level of compactness or homogeneity of data members in each cluster. This value obtained from the average distance of all data against each cluster centroid.
The smaller Average Within Centroid Distance value, then the more compact clusters are formed, which means data members in a cluster have a high level of similarity. This value shows how much meeting data distribution against cluster center.
In the research this metric This is used as indicator supporters For evaluating how much homogeneous data quality food in each cluster formed.

**2.4.3 Evaluation Criteria**

Determination of the optimal number of clusters is carried out with compare results evaluation on several scenario number of clusters, namely K=3 and K=5, with criteria as following :

1. Smallest DBI value shows quality best cluster separation.
2. The smallest Average Within Centroid Distance value shows the best level of cluster cohesiveness.
3. If it occurs difference results between metrics, then optimal cluster selection is carried out based on trade-off considerations between cluster separation and cluster compactness.

Through evaluation use second metric said, research This can determine the most representative number of clusters For describe pattern grouping quality MBG Program food in general objective and measurable.
Study related data grouping using K-Means algorithm has Lots applied to various domains, including grouping food consumption patterns and evaluating product quality [10], [11]. In addition, some studies also show that use cluster validation is very important to ensure quality results for optimal clustering [12], [13]. Utilization clustering method on large multidimensional datasets has also been proven capable of providing effective data segmentation results across multiple field applications [14]. In the context of implementation of the MBG program, data-based food quality evaluation is expected to support transformational improvement in quality service nutrition in a sustainable way [15]. A similar approach has also been applied to research educational data management using the K-Means method for clustering student performance as a form of implementation of clustering in data-based decision making [16].

## 3. RESULTS AND DISCUSSION

### 3.1 Data Description

The following data set components are used in this study:

Table 2: Dataset Component Table
[Source, Author, 2026]

| No | Attribute Name | Data Type |
|---|---|---|
| 1 | ID_Menu | Integer / ID |
| 2 | Menu_Name | String / Nominal |
| 3 | Calories | Numeric |
| 4 | Protein_g | Numeric |
| 5 | Fat_g | Numeric |
| 6 | Carbohydrate_g | Numeric |
| 7 | Freshness_1_10 | Integer |
| 8 | Cleanliness_1_10 | Integer |
| 9 | Serving Temperature_C | Numeric |
| 10 | Eligibility_1_10 | Integer |

Clustering test results on quality data MBG Food program shows comparison performance as follows:

Table 3: Comparison Table of Results
[Source, Author, 2026]

| K | Avg. Within Centroid Distance | Davies-Bouldin Index |
|---|---|---|
| 3 | 0.617 | 0.214 |
| 5 | 0.578 | 0.224 |

The clustering process in this study was implemented using RapidMiner, as illustrated in Figure X. The process consists of several main stages: data input, preprocessing, clustering, and performance evaluation.

The process begins with the Read Excel operator, which is used to import the dataset containing MBG food quality attributes. Next, the data undergoes a normalization process using Z-Transformation to ensure that all attributes are on the same scale and contribute proportionally to the clustering process.

After preprocessing, the data is processed using the K-Means clustering algorithm, which groups the data based on similarity between attributes. Finally, the clustering results are evaluated using the Performance operator, which calculates cluster quality metrics such as the Davies-Bouldin Index (DBI) and Average Within Centroid Distance.

The DBI value at K=3 is lower compared to K=5, showing that K=3 has better cluster separation quality [9]. However, the Average Within Centroid Distance value at K=5 is larger smaller, indicating that the cluster is more compact. Although thus, considering the trade-off between separation and compactness, K=3 is selected as

optimal number of clusters because it produces the best balance in clustering quality. The clustering results are then interpreted into three categories of quality MBG Program food, namely quality high quality medium, and quality low. Grouping This can help party managers evaluate and improve.

### 3.2 Discussion

The comparison between clustering scenarios shows that each configuration provides different characteristics. The K=3 model demonstrates better cluster separation, as indicated by its lower Davies-Bouldin Index (DBI), while the K=5 model produces more compact clusters, reflected in a lower Average Within Centroid Distance. This indicates a trade-off between cluster separation and compactness, where improving one aspect does not necessarily improve overall clustering quality.

The following figures present the visualization of clustering results for K=3 and K=5

a. K= 3


Figure 1. Clustering Visualization for K=3
[Source, Author, 2026]

b. K= 5

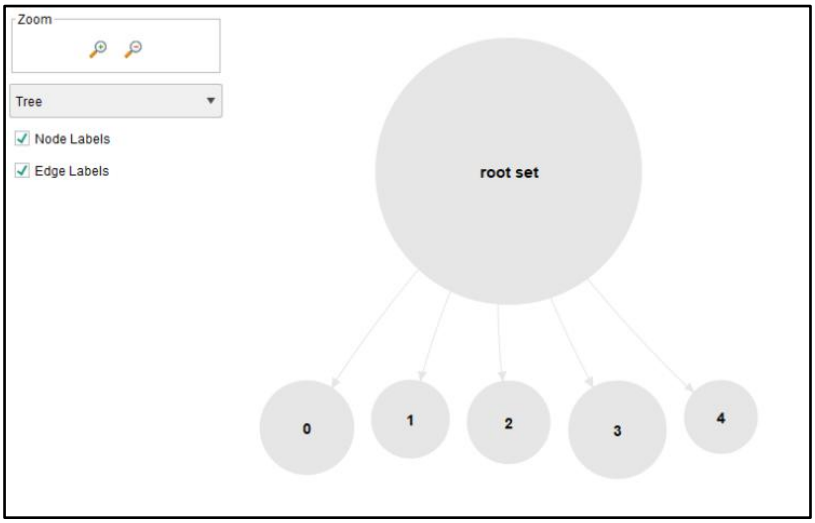


Figure 2. Clustering Visualization for K=5
[Source, Author, 2026]

The results suggest that relying solely on cluster compactness can be misleading, as well-separated clusters are more important for ensuring clear differentiation between data groups. In this context, the K=3 configuration provides a more balanced clustering structure, where the separation between clusters is more distinct and meaningful.

This finding is further supported by the clustering visualization. In the K=3 graph, the data is clearly divided into three main clusters originating from the root set, forming a simple and well-structured grouping pattern. The separation between clusters appears more distinct, indicating that each group represents significantly different characteristics of food quality. In contrast, the K=5 graph shows a more fragmented structure, where the data is divided into five smaller clusters. Although this configuration provides more detailed segmentation, the cluster boundaries become less distinct and more difficult to interpret, indicating potential over-segmentation.

The use of multiple attributes, including nutritional and qualitative factors, plays a significant role in improving clustering performance. By incorporating diverse variables such as Calories, Protein, Fat, Carbohydrates, Freshness, Cleanliness, Serving Temperature, and Eligibility, the model is able to capture more complex patterns in food quality data. This multi-attribute approach enhances the representativeness of the clustering results compared to single-attribute methods.

From an application perspective, the clustering results are easier to interpret when fewer but meaningful clusters are formed. The K=3 model produces a simpler and more practical structure, which supports its use in real-world scenarios such as monitoring and evaluating food quality in the MBG program.

## 4. CONCLUSION

Based on the results of this study, the implementation of the K-Means algorithm has been proven effective in clustering food quality in the Free Nutritious Meal Program (MBG) using all available attributes. This data mining approach provides a more objective, systematic, and data-driven method for evaluating food quality compared to conventional manual assessment.

The clustering evaluation results show that in the K=3 scenario, the data distribution consists of 160 items in cluster 0, 156 items in cluster 1, and 184 items in cluster 2, with a Davies-Bouldin Index (DBI) value of 0.214 and an Average Within Centroid Distance of 0.617. Meanwhile, in the K=5 scenario, the data is more fragmented into five clusters (111, 91, 97, 116, and 85 items), with a DBI value of 0.224 and a lower Average Within Centroid Distance of 0.578, indicating higher cluster compactness.

From a comparative perspective, although the K=5 model produces more compact clusters (lower within-centroid distance), it results in a higher DBI value, which indicates weaker separation between clusters. In contrast, the K=3 model achieves a lower DBI value, demonstrating better inter-cluster separation and clearer differentiation among data groups. This indicates that the K=3 configuration provides a more optimal clustering structure, avoiding over-segmentation while maintaining meaningful group distinctions.

Therefore, it can be concluded that the optimal number of clusters is K=3, as it provides the best balance between cluster separation and interpretability, even though K=5 offers slightly better compactness. The K=3 model is more suitable for practical implementation, as it classifies MBG food quality into three meaningful categories: high, medium, and low, which aligns well with decision-making needs in program evaluation.

This study contributes scientifically by demonstrating that cluster validation using the Davies-Bouldin Index is essential in determining the optimal number of clusters, rather than relying solely on cluster compactness metrics. The findings also highlight that increasing the number of clusters does not necessarily improve clustering quality, emphasizing the importance of balancing separation and cohesion in clustering evaluation.

In practical terms, the results of this study can support program managers in conducting more effective, consistent, and data-driven food quality monitoring. For future research, it is recommended to compare the K-Means algorithm with other clustering approaches such as Hierarchical Clustering, DBSCAN, and Fuzzy C-Means, as well as to incorporate larger datasets and more diverse attributes to enhance accuracy and generalizability.

## REFERENCES


[1] J. Han, M. Kamber, and J. Pei, Data Mining: Concepts and Techniques, 3rd ed. San Francisco, CA, USA: Morgan Kaufmann, 2012.

[2] P.-N. Tan, M. Steinbach, and V. Kumar, Introduction to Data Mining. Boston, MA, USA: Pearson, 2019.

[3] M. I. Faidah and Z. Fatah, "K-Means Clustering With RapidMiner For Identification of Bestseller Products," JUSIFOR, vol. 4, no. 1, 2025.

[4] M. Noviati et al., "Clustering Sales Data Product Food at the Yogya Siliwangi Department Store with the K-Means Method," MEANS, vol. 7, no. 2, 2022.

[5] G. Triyandana, L. A. Putri, and Y. Umaidah, "Application of Data Mining for Food and Beverage Menu Grouping Based on Sales Level Using the K-Means Method," Journal of Applied Informatics and Computing, vol. 6, no. 2, 2022.

[6] F. Nurulhikmah and D. N. E. Abdi, "Classification of Foods Based on Nutritional Content Using K-Means and DBSCAN Clustering Methods," Teknika, vol. 13, no. 3, 2024.

[7] S. E. N. Putra and M. Furqan, "Application of Data Mining in Grouping Product Quality Using K-Means Clustering Algorithm," CESS, vol. 9, no. 1, 2024.
[8] M. A. Irwansyah et al., "K-Means Analysis with RapidMiner for Classification 'The Quality of Elementary School Education in Indonesia'," ZONAsi, vol. 7, no. 2, 2025.
[9] C. Sinaga and S. P. Sipayung, "Implementation of K-Means Clustering with Davies-Bouldin Index Evaluation," KAKIFIKOM, vol. 7, no. 2, 2026.
[10] I. T. Julianto et al., "Data Mining Clustering Food Expenditure in Indonesia," Jurnal Teknik Informatika (JUTIF), vol. 3, no. 6, 2022.
[11] S. N. Adzra, F. N. Hasan, and A. Y. Kuntoro, "Application of Data Mining in Student Academic Performance Assessment With the K-Means Clustering Method," Journal of Scientific Informatics, vol. 13, no. 2, 2025.
[12] Y. A. Joarder and M. Ahmed, "A Hybrid Algorithm Based Robust Big Data Clustering," arXiv preprint arXiv:2002.09380, 2020.
[13] B. Venkateswarlu and G. S. V. P. Raju, "Mine Blood Donors Information through Improved K-Means Clustering," arXiv preprint arXiv:1309.2597, 2013.
[14] L. W. Yerbury et al., "On the Use of Relative Validity Indices for Comparing Clustering Approaches," arXiv preprint arXiv:2404.10351, 2024.
[15] Albaburrahim et al., "Free Nutritional Meal Program: An Analysis Critical Transformation of Indonesian Education Towards Golden Generation 2045," Entita, vol. 1, no. 2, 2025.
[16] F. Fitriasih et al., "Training Data Management Using the K-Means Method For Clustering Student Achievement," Polytechnic Purbaya, Tegal, Tech. Rep., 2026.