Clustering Student Satisfaction Levels Regarding the Free Nutritious Food Program at SMAN 1 Delitua Using the K-Means Algorithm

Sujatmiko Ginting, Muhammad Syahputra Novelan

ABSTRACT

The Free Nutritious Food (MBG) Program is a strategic policy initiated by the Indonesian government to improve the nutritional status and academic performance of students. This study aims to cluster the satisfaction levels of students at SMAN 1 Delitua regarding the MBG Program using a data mining approach with the K-Means clustering algorithm. A quantitative approach was employed, utilizing a Likert-scale questionnaire (1–5) to evaluate eight satisfaction indicators: taste, quality, menu variation, portion, cleanliness, timeliness, health benefits, and overall satisfaction. The study population comprised all students at SMAN 1 Delitua, with a sample of 600 respondents (350 males and 250 females) selected via stratified random sampling. The optimal number of clusters was determined using the Elbow Method and the Silhouette Score. The K-Means algorithm identified five distinct clusters representing varying satisfaction levels: very low (6.33%), low (13.00%), moderate (62.67%), high (11.17%), and very high (6.83%). The predominance of moderate to high satisfaction levels indicates that the MBG Program is generally well-received by the students. This research contributes a data-driven evaluation model for school nutrition policies and offers targeted recommendations for improving satisfaction among underperforming segments.

***Keywords**: K-Means Clustering, student satisfaction, Free Nutritious Food Program, data mining, program evaluation*

Introduction

Sujatmiko Ginting
Magister of Information Technology, Universitas Pembangunan Panca Budi, Indonesia
e-mail: sujatmikoginting5@gmail.com

Muhammad Syahputra Novelan
e-mail: putranovelan@dosen.pancabudi.ac.id
2nd International Cofference on the Epicentrum of Economic Global Framework (ICEEGLOF)
Theme: Navigating The Future: Business and Social Paradigms in a Transformative Era.
https://proceeding.pancabudi.ac.id/index.php/ICEEGLOF

Adequate balanced nutrition is an essential foundation for the physical growth and cognitive development of students, particularly at the senior high school level. Micronutrient and macronutrient deficiencies during adolescence can negatively impact learning concentration, immune system resilience, and academic achievement. To address this issue, the Indonesian government initiated the Free Nutritious Food (MBG) Program, targeting students across various educational levels.

The effectiveness of the MBG Program is measured not only by food distribution metrics but also by the satisfaction levels of the student beneficiaries. Satisfaction is a key indicator of service quality, reflecting the alignment between student expectations and the reality of the service provided. Student satisfaction data is multidimensional and heterogeneous; individual students differ in their assessments of aspects such as taste, portion size, menu variation, and food cleanliness. This variation forms a complex data pattern that is difficult to interpret in aggregate.

Data mining, specifically clustering analysis, offers a robust solution for identifying natural groupings within student satisfaction data. Clustering is an unsupervised learning method that groups data based on characteristic similarities without requiring prior class labels. The K-Means clustering algorithm was selected for this study due to its efficiency in processing large-scale numerical data, ease of implementation, and interpretability. Novelan et al. (2023) asserted that machine learning approaches, specifically K-Means, are highly effective in resolving imbalanced and large-scale data classification problems, such as those found in vehicle routing and goods distribution. Furthermore, Novelan et al. (2025) demonstrated that K-Means is reliable for clustering healthcare service data based on patient visit patterns to enhance operational efficiency and patient satisfaction, which shares contextual similarities with the evaluation of school nutrition programs.

A clustering analysis of student satisfaction regarding the MBG Program has not previously been conducted at SMAN 1 Delitua. This study serves as a data-driven preliminary evaluation to support continuous program quality improvement. The primary research gap addressed is the lack of studies utilizing data mining to segment student satisfaction in local school nutrition programs, specifically within the Delitua region. This research contributes by profiling student groups based on satisfaction levels, providing an empirical basis for decision-making by school administrators and program managers. This aligns with the recommendations of Novelan et al. (2025), which emphasize the importance of data-driven analysis to support strategic decision-making in the public service sector.

Literature Review

The MBG Program is a government policy designed to mitigate nutritional deficits and enhance the quality of human resources. The program targets school-aged children from primary to secondary levels by providing free, balanced, and nutritious Foods (Kementerian Kesehatan RI, 2023). The success of this program requires continuous monitoring and evaluation, primarily through measuring beneficiary satisfaction (World Food Programme, 2020). In the context of public services like nutrition programs, measuring student satisfaction involves assessing various service attributes, including food quality, cleanliness, and service timeliness. The standard instrument used is a questionnaire employing a Likert Scale (1 = strongly dissatisfied to 5 = strongly satisfied); while this yields ordinal data, it is frequently treated as interval data in multivariate statistical analyses, such as clustering, given specific methodological considerations (Sullivan & Artino, 2013).

Clustering is a fundamental data mining technique classified under unsupervised learning. Its objective is to partition a dataset into groups (clusters) that are internally homogeneous and externally heterogeneous. K-Means is a widely used partitional clustering algorithm that

operates by partitioning n objects into K clusters, with each object assigned to the cluster featuring the nearest mean or centroid (MacQueen, 1967). K-Means is computationally efficient and highly suitable for medium to large numerical datasets (Jain, 2010). In their research on Vehicle Routing Problem (VRP) optimization, Novelan et al. (2023) demonstrated that K-Means effectively handles imbalanced data with high accuracy (81.86% at K=15), highlighting the algorithm's cross-domain flexibility. Additionally, an analysis of patient medical records utilizing K-Means successfully identified patient segments based on visit times to support adaptive service strategies (Novelan et al., 2025). The implementation of K-Means has also proven effective in classifying organic and inorganic waste by determining the optimal number of clusters using the Silhouette Score (Asyifa & Novelan, 2025).

Research Methodology

3.1 Research Type and Approach

This study employed a quantitative approach utilizing descriptive and exploratory methods. A quantitative approach was selected to numerically measure and cluster student satisfaction levels using survey data. The descriptive method was used to outline respondent characteristics and the distribution of responses across satisfaction indicators. The exploratory method was applied via the K-Means clustering algorithm to identify previously unknown patterns in student satisfaction.

3.2 Data Processing Stages

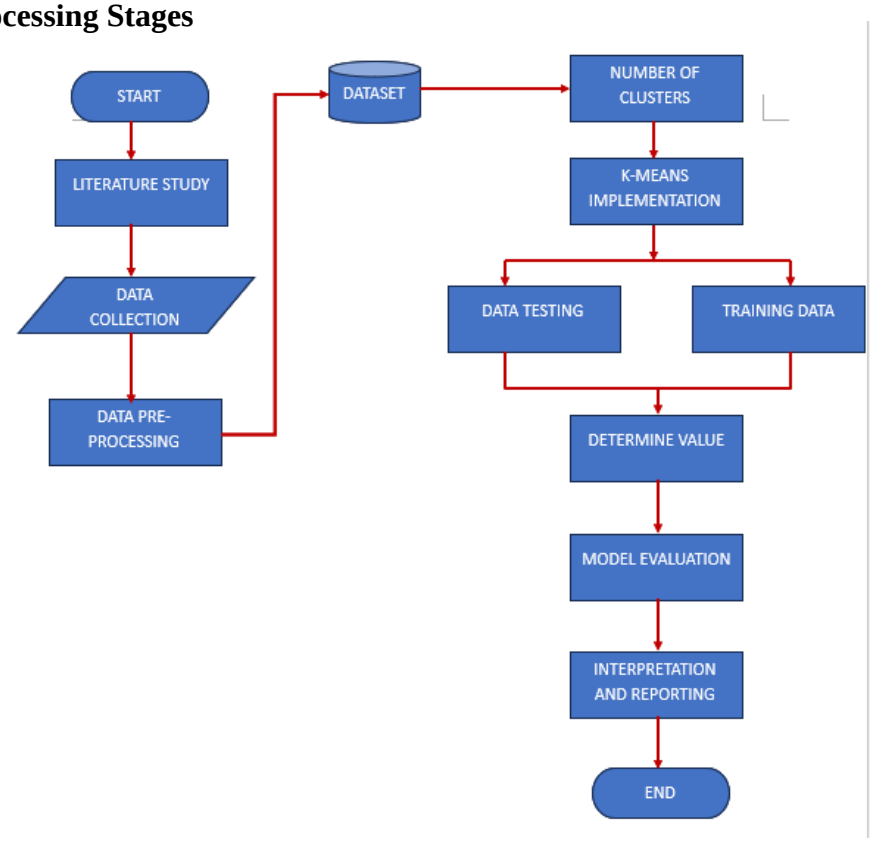


Figure 1. Research Flow

Result

4.1 Literature Study

This stage was conducted to review theories relevant to the MBG program, student satisfaction concepts, and the K-Means algorithm. The literature review served as the theoretical foundation for developing research instruments and interpreting the analytical results.

4.2 Data Collection

Data were collected via online questionnaires distributed to the target respondents from March 5, 2026, to August 5, 2026. Table 1 presents the demographic characteristics of the 600 respondents based on gender.

Table 1. Respondent Characteristics Based on Gender

Gender	Total	Percentage
Male	350	58.33%
Female	250	41.67%
Total	600	100%

4.3 Data Pre-processing

The dataset underwent a cleaning process to address missing values and outliers. Subsequently, the data were normalized using the Min-Max Scaling method to standardize the scale across all variables.

Table 2. Descriptive Statistics of Satisfaction Indicators

Indicator	Mean	Median	Mode	Std. Deviation	Min	Max
Taste	3.07	3	3	0.63	1	5
Quality	3.07	3	3	0.63	1	5
Variation	3.72	4	4	0.73	1	5
Portion	4.37	5	5	1.06	1	5
Cleanliness	3.72	4	4	0.73	1	5
Timeliness	4.37	5	5	1.06	1	5
Benefits	3.72	4	4	0.73	1	5
Satisfaction	3.72	4	4	0.73	1	5

4.4 Determination of the Optimal Number of Clusters

The optimal number of clusters was determined using the Elbow Method and the Silhouette Coefficient. The Elbow Method analyzes the Within-Cluster Sum of Squares (WCSS) for various K values, while the Silhouette Coefficient measures intra-cluster cohesion against inter-cluster separation. These evaluations were conducted for the range of K = 2 to K = 10.

Table 3. Elbow Method and Silhouette Score Results

K	WCSS	Silhouette Score
2	3384.8320	0.5588

3	2347.2575	0.5466
4	2098.0683	0.4358
5	1937.4932	0.4394
6	1804.9876	0.4460
7	1699.0041	0.4436
8	1612.3243	0.4673
9	1539.8524	0.4702
10	1473.9311	0.4760

The WCSS values exhibit a steep decline from K=2 (3384.84) to K=3 (2347.26) and continue to decrease steadily until K=5 (1937.49). Beyond K=5, the rate of WCSS reduction slows significantly, indicating diminishing returns. This suggests that adding clusters beyond K=5 does not yield a substantial reduction in internal variance. While the highest Silhouette Scores occur at K=2 (0.5588) and K=3 (0.5466), dividing the dataset into only two or three clusters often results in under-segmentation, failing to provide detailed, actionable insights. Therefore, K=5 was selected as the optimal balance point to maintain meaningful segment complexity, low intra-cluster variance, and adequate cluster separation.

4.5 Implementation of the K-Means Algorithm

The K-Means algorithm utilized K-Means++ centroid initialization to enhance the stability of the results. The iterative process continued until convergence was achieved. Following the selection of K = 5, the algorithm was applied to the dataset.

Table 4. Final Centroid Values of Each Cluster

Indicator	Very Low	Low	Moderate	High	Very High
Taste	1.00	2.00	3.00	3.00	5.00
Quality	1.00	2.00	3.00	3.00	5.00
Variation	1.00	2.00	3.00	4.00	5.00
Portion	1.00	2.00	3.00	5.00	5.00
Cleanliness	1.00	2.00	3.00	4.00	5.00
Timeliness	1.00	2.00	3.00	5.00	5.00
Benefits	1.00	2.00	3.00	4.00	5.00
Satisfaction	1.00	2.00	3.00	4.00	5.00

Table 5. Total and Percentage of Members in Each Cluster

Cluster	Number of Members	Percentage
Very Low	38	6.33%
Low	78	13.00%
Moderate	376	62.67%
High	67	11.17%
Very High	41	6.83%
Total	600	100%

4.6 Model Evaluation

Cluster quality was assessed using the Silhouette Coefficient and the Davies-Bouldin Index. The Silhouette Coefficient measures the similarity of an object to its own cluster relative to other clusters, with values ranging from -1 to 1. The Davies-Bouldin Index evaluates the ratio of inter-cluster distance to intra-cluster distance, where lower values indicate superior cluster quality.

4.7 Results and Cluster Interpretation

The clustering results were interpreted based on the centroid values across all satisfaction dimensions. The characteristics of each cluster are detailed below to facilitate targeted policy recommendations:

- Cluster 1: Very Low Satisfaction Level (6.33%). Comprising 38 students, this cluster exhibits average scores of 1.0–1.5 across all satisfaction indicators. These students expressed extreme dissatisfaction with the taste, quality, variation, and overall execution of the MBG Program. This group requires immediate attention from program managers to diagnose and address the root causes of their dissatisfaction.
- Cluster 2: Low Satisfaction Level (13.00%). Consisting of 78 students, this group shows average satisfaction scores between 2.0 and 2.5. Students in this segment lean toward dissatisfaction. The lowest-scoring indicator in this cluster is 2.00, identifying a specific priority area for targeted improvements.
- Cluster 3: Moderate Satisfaction Level (62.67%). As the largest group with 376 students, average scores range from 3.0 to 3.5. These students report adequate satisfaction with the MBG Program, though there remains discernible room for enhancement.
- Cluster 4: High Satisfaction Level (11.17%). Comprising 67 students, this sizable segment reported average scores between 4.0 and 4.5. The majority of students in this cluster are highly satisfied, indicating that the program operates efficiently and meets their expectations.
- Cluster 5: Very High Satisfaction Level (6.83%). Consisting of 41 students, this cluster features average scores of 4.5–5.0. These students provided exceptional ratings across all program dimensions, serving as the benchmark for successful MBG Program implementation.

Conclusion

Based on the research findings and discussion, this study concludes the following:

1. The application of the K-Means clustering algorithm to the satisfaction data of 600 students at SMAN 1 Delitua regarding the Free Nutritious Food Program successfully yielded five distinct clusters. The optimal cluster count (K = 5) was empirically validated using the Elbow Method and Silhouette Score.
2. The respondent distribution across the clusters is as follows: Cluster 1 (Very Low Satisfaction) at 6.33%, Cluster 2 (Low Satisfaction) at 13.00%, Cluster 3 (Moderate Satisfaction) at 62.67%, Cluster 4 (High Satisfaction) at 11.17%, and Cluster 5 (Very High Satisfaction) at 6.83%.
3. The majority of the student population (62.67%) falls into the moderate satisfaction level, signifying that the MBG Program at SMAN 1 Delitua is generally effective and meets most student expectations. However, the 13.00% of students with low satisfaction and the 6.33% with very low satisfaction highlight segments that require targeted programmatic interventions.
4. The most significant differentiating indicators among the clusters are taste, quality, and menu variation. These aspects should be prioritized for future program enhancements.

References

[1] Novelan, M. S., Efendi, S., Sihombing, P., & Mawengkang, H. (2023). Vehicle routing problem optimization with machine learning in imbalanced classification vehicle route data. Eastern-European Journal of Enterprise Technologies, 5(3(125)), 49–56. https://doi.org/10.15587/17294061.2023.288280

[2] Asyifa, N., & Novelan, M. S. (2025). Performance Analysis of the K-Means Algorithm in Classifying Organic and Inorganic Waste Types. Proceedings of International Conference on Islamic Community Studies, Universitas Pembangunan Panca Budi.

[3] Padilla, E. M., Safii, A., Nasution, I., Anugrah, M. F., Harahap, N. S., & Novelan, M. S. (2025). Analysis of Patient Medical Records Using the K-Means Clustering Algorithm Based on Visit Time as a Service Strategy Approach. Jurnal Bisantara Informatika, 9(1). http://bisantara.amikparbinanusantara.ac.id/index.php/bisantara/article/view/161

[4] Fauzi, A. A., Lestanti, S., & Wulansari, Z. (2023). Pengelompokkan Kepuasan Siswa Terhadap Pembelajaran Daring Menggunakan Algoritma K-Medoids. Jurnal Informatika Polinema, 9(3), 307–314. https://doi.org/10.33795/jip.v9i3.1255

[5] Kusuma Al Atros, K. S. H., Padri, A. R., Nurdiawan, O., Faqih, A., & Anwar, S. (2021). Model Klasifikasi Analisis Kepuasan Pengguna Perpustakaan Online Menggunakan K-Means dan Decision Tree. JURIKOM (Jurnal Riset Komputer), 8(6), 323. https://doi.org/10.30865/jurikom.v8i6.3680

[6] MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations. Proceedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability, 1, 281-297.

[7] Mendrofa, T. R., & [Author]. (2026). Analisis Tingkat Kepuasan Siswa terhadap Program Makan Bergizi Gratis di SMA Negeri 1 Manokwari dengan Pendekatan Mining. Jurnal Indonesia: Manajemen Informatika dan Komunikasi, 7(2). https://doi.org/10.63447/jimik.v7i2.1842

[8] Rahmi, N., Baisa, L. Y., & Sumendap, A. L. (2026). Analysis of Students' Perceptions of the Free Nutritious Food Program (MBG) Based on K-Means Clustering. Indonesian Journal of Artificial Intelligence and Data Mining, 9(1). http://dx.doi.org/10.24014/ijaidm.v9i1.39240

[9] Saputra, V., Rifaldi, D., Ramadhan, F. P., & Mulyadi, I. P. (2026). Implementasi K-Means Mengukur Kepuasan Siswa Terhadap Pelayanan pada SMKS Pelita Raya Jambi. Jurnal Software Engineering and Information System (SEIS), 6(1). https://doi.org/10.37859/seis.v6i1.9636

[10] Suryanata, M. G., Pane, D. H., & [Author]. (2019). Implementasi Algoritma K-Means Untuk Mengukur Tingkat Kepuasan Siswa Terhadap Proses Pembelajaran.

Appendix

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.cluster import KMeans
from sklearn.preprocessing import StandardScaler
from sklearn.decomposition import PCA
from sklearn.metrics import silhouette_score

# ====================================================
```

```
urutan = rata2_cluster.sort_values().index

label_kategori = ['Sangat Rendah','Rendah','Sedang','Tinggi','Sangat Tinggi']
mapping = {old: new for old, new in zip(urutan, label_kategori)}
df['Cluster'] = df['Cluster_raw'].map(mapping)

# Hitung jumlah & persentase
hasil = df['Cluster'].value_counts().reindex(label_kategori)
persen = (hasil / len(df) * 100).round(2)

print("="*60)
print("1. HASIL PENGELOMPOKAN KEPUASAN SISWA (K = 5)")
print("="*60)
rekap = pd.DataFrame({'Jumlah Siswa': hasil,
                      'Persentase (%)': persen})
rekap.loc['TOTAL'] = [hasil.sum(), 100.00]
print(rekap)

# Rata-rata indikator per cluster (opsional, biar informatif)
print("\nRata-rata skor tiap indikator per cluster:")
print(df.groupby('Cluster')[fitur].mean().round(2).reindex(label_kategori))

# ========================================================
# 3. SCATTER PLOT HASIL CLUSTERING (PCA 2D)
# ========================================================
pca   = PCA(n_components=2)
X_pca = pca.fit_transform(X_scaled)

plt.figure(figsize=(10, 7))
warna = {'Sangat Rendah':'#d62728','Rendah':'#ff7f0e',
         'Sedang':'#bcbd22','Tinggi':'#2ca02c',
         'Sangat Tinggi':'#1f77b4'}
for lbl in label_kategori:
    mask = df['Cluster'] == lbl
```