

Application of Data Mining Algorithms to Determine Priority Schools for the Free Nutritious Meal Program Based on Social Indicators in Pinrang Regency

Anas Nurdin Hadi^{1*}, Jabaluddin Hamud¹, Siti Rahmah Sindar¹, Ahmad Husain¹, Azwar H.²

¹Department Sains, Institut Teknologi Bacharuddin Jusuf Habibie, Indonesia

²Department of Informatics, Universitas Ichsan Sidenreng Rappang, Indonesia

anas@ith.ac.id

ABSTRACT

Article History:

Received : 12-06-2026

Revised : 05-07-2026

Accepted : 07-07-2026

Online : 20-10-2026

Keywords:

Data Mining;

K-Means Clustering;

Social Indicators;

Silhouette Method;

Priority School.



This study aims to determine priority schools for the Free Nutritious Meal Program in Pinrang Regency based on students' social indicators using a data mining approach. The implementation of educational assistance programs requires objective and data-driven decision-making to ensure that support is distributed to schools with the highest level of social vulnerability. Therefore, this research applies the K-Means Clustering algorithm to classify schools according to the socioeconomic conditions of their students. The research employed a quantitative approach involving 593 students from 10 senior high schools and vocational schools in Pinrang Regency. The dataset was collected using questionnaires and school administrative records. Students were evaluated using multidimensional socioeconomic indicators representing household economic conditions, educational background, housing characteristics, family dependency, transportation accessibility, and social assistance status. Each indicator was transformed into a numerical score ranging from 1 to 4 to quantify the level of socioeconomic vulnerability. The clustering process was conducted using the K-Means Clustering algorithm, while the optimal number of clusters was determined using the Silhouette Method. The results showed that the optimal clustering structure consisted of four clusters with a silhouette score of 0.234. Cluster 1 represented the highly prioritized group with the highest socioeconomic vulnerability, while Cluster 4 represented the priority group, Cluster 3 represented the moderately prioritized group, and Cluster 2 represented the non-priority group. The analysis indicated that School B and School H were categorized as highly prioritized schools because most of their students belonged to Cluster 1. Meanwhile, School G was categorized as a priority school, and School D was classified as moderately prioritized. The findings demonstrate that the K-Means Clustering method can effectively support educational policy decision-making by identifying school priority categories based on multidimensional social indicators. The novelty of this study lies in the development of a data-driven school prioritization framework that integrates ten multidimensional socioeconomic indicators into a K-Means Clustering model. Unlike the current prioritization approach, which primarily considers general vulnerability criteria and regional classification, the proposed framework quantitatively evaluates students' socioeconomic conditions to support a more objective, transparent, and evidence-based prioritization of schools for the Free Nutritious Meal Program.



<https://doi.org/10.31764/jtam.v10i4.40194>



This is an open access article under the [CC-BY-SA](#) license

A. INTRODUCTION

The rapid advancement of information technology and digital transformation has significantly transformed decision-making processes across various sectors, including education. Educational institutions increasingly employ data-driven approaches to support planning, resource allocation, institutional evaluation, and evidence-based policy formulation. In this context, Educational Data Mining (EDM) has emerged as an interdisciplinary field that integrates data mining, machine learning, and learning analytics to discover meaningful patterns from educational data and improve educational decision-making. Recent studies have demonstrated that EDM contributes not only to predicting student performance but also to optimizing institutional planning, educational management, and policy development through intelligent data analysis (Aldowah et al., 2019; Zaki & Wagner, 2020).

The increasing availability of educational and socioeconomic data offers governments valuable opportunities to develop objective and transparent mechanisms for identifying schools and students requiring priority intervention. Rather than relying solely on administrative assessments, data-driven decision-making enables policymakers to integrate multiple socioeconomic indicators simultaneously, thereby improving the accuracy, transparency, and fairness of beneficiary identification (Ifenthaler et al., 2023; Romero & Ventura, 2020).

School feeding programs have been widely recognized as one of the most effective interventions for improving children's nutritional status, educational participation, and learning outcomes. Recent empirical studies have consistently shown positive impacts on school attendance, cognitive development, dietary diversity, and overall student well-being, particularly in low- and middle-income countries (Barnabas et al., 2024; Wang et al., 2021; Enchikova et al., 2024). Furthermore, systematic reviews have confirmed that school meal programs contribute to improvements in anthropometric indicators, learning participation, and educational retention, although their effectiveness strongly depends on appropriate beneficiary targeting mechanisms (Kristjansson et al., 2022).

Despite the importance of accurate targeting, the identification of priority schools for educational assistance programs is frequently conducted using conventional administrative approaches that rely heavily on subjective assessments and limited socioeconomic indicators. Such approaches may not adequately represent the multidimensional characteristics of students' socioeconomic vulnerability. Recent evidence suggests that socioeconomic status should be understood as a multidimensional construct encompassing household income, parental education, occupational background, home resources, and broader social and environmental conditions that collectively influence students' educational opportunities and learning outcomes (Tan, 2024). Furthermore, systematic reviews of educational equity demonstrate that disparities in educational access continue to be strongly associated with socioeconomic status, geographical location, family background, and other intersecting social determinants, indicating the necessity of adopting multidimensional approaches when determining priority beneficiaries of educational assistance programs (Enchikova et al., 2024; Fitzgerald et al., 2025). Likewise, multidimensional socioeconomic vulnerability indices have been recommended as more reliable instruments for identifying disadvantaged populations

because they integrate economic, educational, housing, and social deprivation indicators rather than relying solely on income-based measures (Palermos et al., 2024).

Social vulnerability refers to the susceptibility of individuals or communities to adverse socioeconomic conditions that limit access to education, health services, and other essential resources. Contemporary studies indicate that vulnerability is multidimensional and influenced by household income, employment, educational attainment, housing quality, family dependency burden, and accessibility to public services (Mah et al., 2023; Ajtai et al., 2023; Majumder et al., 2023). Therefore, multidimensional vulnerability assessment has become an important approach for developing more accurate and equitable social assistance policies.

Data mining has become an essential analytical approach for extracting valuable knowledge from large-scale and multidimensional datasets, enabling researchers and policymakers to support evidence-based decision-making across various application domains. Recent advances in data mining and big data analytics have significantly enhanced the capability to identify hidden structures and meaningful patterns from complex datasets, particularly in socioeconomic analysis, public governance, and evidence-based policymaking (Hossin et al., 2023; Abuljadail et al., 2023). Among various data mining techniques, clustering is recognized as one of the most effective unsupervised learning methods because it groups observations according to their intrinsic similarities without requiring predefined class labels. This capability allows clustering algorithms to reveal natural data structures that are difficult to detect using conventional analytical approaches (Abuljadail et al., 2023; Wang et al., 2021). Consequently, clustering techniques have been increasingly employed to analyze multidimensional socioeconomic characteristics, identify vulnerable population groups, and support more objective and data-driven policy formulation. By grouping observations with similar characteristics, clustering facilitates the identification of hidden patterns in complex socioeconomic data, thereby improving the targeting of interventions and evidence-based policymaking (Bittencourt et al., 2024; Zhang et al., 2024).

Clustering techniques provide powerful unsupervised learning mechanisms for identifying naturally occurring groups within multidimensional datasets without predefined class labels. These techniques have been widely implemented in educational data mining, public policy analysis, healthcare, and social vulnerability assessment because they enable researchers to discover hidden structures in complex datasets and support evidence-based decision-making (Atif et al., 2023; Chaudhry et al., 2023). Among various clustering algorithms, K-Means remains one of the most popular approaches because of its computational efficiency, scalability, and ease of interpretation.

Previous studies have successfully implemented clustering techniques in educational and social assistance contexts. In Educational Data Mining (EDM), clustering has been widely employed to identify student learning profiles, predict academic performance, detect at-risk students, support scholarship selection, and develop personalized learning recommendation systems (Ifenthaler et al., 2023; Lu et al., 2025). Other studies have applied clustering methods to classify vulnerable populations and support social assistance targeting (Follett & Henderson, 2023; Mah et al., 2023). Nevertheless, limited research has specifically investigated the application of clustering techniques for determining priority schools in the implementation of

large-scale school feeding programs using multidimensional socioeconomic indicators at the school level.

The current implementation of the Free Nutritious Meal Program prioritizes schools based on students' nutritional needs, socioeconomic vulnerability, and regional characteristics, particularly in disadvantaged, frontier, and outermost (3T) areas. While these criteria provide an important policy foundation, the prioritization process has not yet systematically integrated multiple socioeconomic indicators into a quantitative decision-support framework. As a result, variations in school prioritization may occur because the relative socioeconomic vulnerability across schools is not assessed using a standardized multidimensional approach. This study addresses this gap by proposing a K-Means Clustering framework that objectively classifies schools based on multidimensional socioeconomic vulnerability profiles.

This study aims to implement the K-Means Clustering algorithm to classify schools in Pinrang Regency based on multidimensional socioeconomic vulnerability indicators and determine priority categories for the Free Nutritious Meal Program. The proposed approach not only identifies schools with the highest concentration of vulnerable students but also provides an objective and reproducible decision-support mechanism using the Silhouette Method for cluster optimization. Consequently, this study contributes both methodologically and practically by supporting transparent, data-driven educational policy formulation and improving the effectiveness of social assistance allocation..

B. METHODS

This study employed a quantitative approach using data mining techniques to determine priority schools for the Free Nutritious Meal Program based on students' social indicators in Pinrang Regency. The research process consisted of data collection, data preprocessing, descriptive analysis, clustering analysis, and interpretation of the clustering results.

1. Data Collection

This study was conducted in Pinrang Regency, South Sulawesi, Indonesia, with the aim of identifying school priority levels for the implementation of the Free Nutritious Meal Program based on students' socioeconomic vulnerability profiles. The study focused on senior high schools and vocational high schools in the regency, as these schools accommodate students from diverse household backgrounds and are relevant targets for social-based educational assistance. A cross-sectional design was employed, in which student socioeconomic data were collected at a single observation period to provide a snapshot of the vulnerability conditions of students across schools in the study area.

The primary unit of analysis in this study was the student, because the clustering analysis was intended to identify patterns of socioeconomic vulnerability at the individual level before being interpreted at the school level (Mah et al., 2023). Although the final objective of the study was to determine school priority categories, the analytical process first grouped students based on similarities in their social and household characteristics. The resulting student-level clusters were subsequently aggregated to characterize the dominant vulnerability profile of each school and to support the classification of schools into priority categories for the Free Nutritious Meal Program.

The data used in this study were obtained from student socioeconomic records collected through school-based administrative data and supporting questionnaire information. These data were selected because they provide direct information on the social and economic conditions of students and their households, which are highly relevant for identifying vulnerability in the context of educational and nutritional support. The final analytical dataset consisted of 593 student records, each containing information on a set of socioeconomic indicators used to represent household welfare and social disadvantages.

The collected data covered ten social indicators, namely parents' monthly income, parents' occupation, parents' highest educational attainment, home ownership status, number of household dependents, transportation access to school, ownership of social assistance cards, orphan status, living conditions, and the number of siblings attending school. These indicators were selected because they represent multiple dimensions of household socioeconomic conditions that may influence students' access to educational and nutritional support. By collecting data at the student level and incorporating various indicators of vulnerability, this study aimed to provide a more comprehensive basis for clustering analysis and for the subsequent mapping of school priorities in Pinrang Regency.

2. Data Preprocessing

This study aimed to determine priority schools for the Free Nutritious Meal Program using a data mining approach. The research implementation included social indicator processing, descriptive analysis, clustering formation, and school priority categorization. Before clustering was performed, the collected student-level social data were prepared to ensure that all indicators could be represented in a consistent numerical format suitable for K-Means analysis. The raw data were obtained from questionnaires and school administrative records and consisted primarily of categorical socioeconomic information, such as parents' monthly income, parents' occupation, educational attainment, home ownership status, transportation access, social assistance ownership, and other household characteristics (see Table 1). Because these variables were measured in different categorical forms and did not share a common numeric scale, a data preparation stage was required to transform them into comparable numerical inputs for clustering.

In this study, data preparation focused on converting each social indicator into an ordinal score ranging from 1 to 4, where a higher score indicated a higher level of socioeconomic vulnerability. The scoring system was designed to preserve the substantive meaning of each indicator while allowing heterogeneous categorical variables to be analysed jointly in a single clustering framework. For example, lower parental income, unstable employment, lower educational attainment, inadequate housing conditions, limited transportation access, and the presence of social assistance cards were assigned higher scores because these conditions reflect greater social disadvantage and a stronger need for educational support. Conversely, more stable socioeconomic conditions were assigned with lower scores. The detailed scoring rules for all ten indicators are presented in Table 1.

The use of a 1–4 scoring range was intended to standardize the direction and interpretation of all indicators before clustering. First, it ensured that all variables were placed on the same ordinal scale, thereby reducing differences in measurement units across indicators. Second, it

provided a simple and interpretable representation of social vulnerability, in which the score increased in line with the severity of disadvantages. Third, the scoring approach facilitated the integration of multiple dimensions of student welfare into a unified clustering input without requiring separate treatment for each categorical variable. This approach is particularly appropriate in the context of social vulnerability assessment, where many indicators are inherently categorical and need to be translated into a common analytical scale before numerical clustering methods can be applied.

After the scoring process, the ten indicator scores for each student constituted the final feature set used in the clustering analysis. Because all variables had already been transformed into the same ordinal range (1–4), no additional normalization or standardization was applied prior to the K-Means procedure. This decision was made because the main purpose of the transformation stage was to ensure comparability across indicators by placing them on a common scale before clustering. Thus, the scored variables were directly used as the input matrix for the subsequent K-Means analysis.

Only complete student records included in the final dataset were analysed in the clustering stage. Therefore, no additional missing-value imputation procedure was applied in this study. The final analytical dataset consisted of 593 student records from senior high schools and vocational high schools in Pinrang Regency, each represented by ten scored social indicators reflecting the students' socioeconomic vulnerability profile.

Table 1. Social Indicators

No	Social Indicators	Category/Response	Score
1	Parents' Monthly Income	< Rp 1.000.000	4
		Rp 1.000.000 – Rp 2.500.000	3
		Rp 2.500.000 – Rp 5.000.000	2
		> Rp 5.000.000	1
2	Parents' Occupation	Unemployed	4
		Laborer/Freelance Worker	3
		Permanent Employee/Civil Servant/Small Entrepreneur	2
		Entrepreneur/Stable Occupation	1
3	Parents' Highest Educational Level	No Schooling	4
		Elementary School	3
		Junior High School	2
		Senior High School or Higher	1
4	Home Ownership Status	Living with Relatives/Others	4
		Rental/Contract House	3
		Self-Owned Simple House	2
		Self-Owned Permanent House	1
5	Number of Dependents in Household	≥ 6 persons	4
		4–5 persons	3
		2–3 persons	2
		1 person	1
6	Transportation Access to School	Walking > 3 km	4
		Walking < 3 km / Bicycle	3
		Public Transportation	2
		Private Vehicle	1

No	Social Indicators	Category/Response	Score
7	Ownership of KIP/KKS/PKH Assistance	Yes	4
		No	1
8	Orphan Status	Double Orphan	4
		Single Orphan	3
		Both Parents Alive	1
9	Living Conditions	Uninhabitable/Semi-permanent House	4
		Simple Livable House	3
		Proper Livable House	2
		Permanent and Comfortable House	1
10	Number of Siblings Attending School	≥ 4 persons	4
		2–3 persons	3
		1 person	2
		None	1

The scoring interpretation used in this study was as follows:

- 25–30 : Highly Needy
- 18–24 : Moderately Needy
- 10–17 : Considered
- 0–9 : Not Priority

3. Descriptive Analysis

The descriptive analysis results were visualized using bar plots for each variable. Such visualizations facilitate the identification of distribution patterns and socioeconomic disparities across respondents before conducting multivariate analyses (Enchikova et al., 2024). The findings showed that many students traveled to school by walking more than 3 kilometers, indicating transportation difficulties. A significant proportion of students also came from families with monthly incomes below Rp 1.000.000, unemployed parents, and inadequate housing conditions. These findings indicate the existence of groups of students with high levels of social vulnerability.

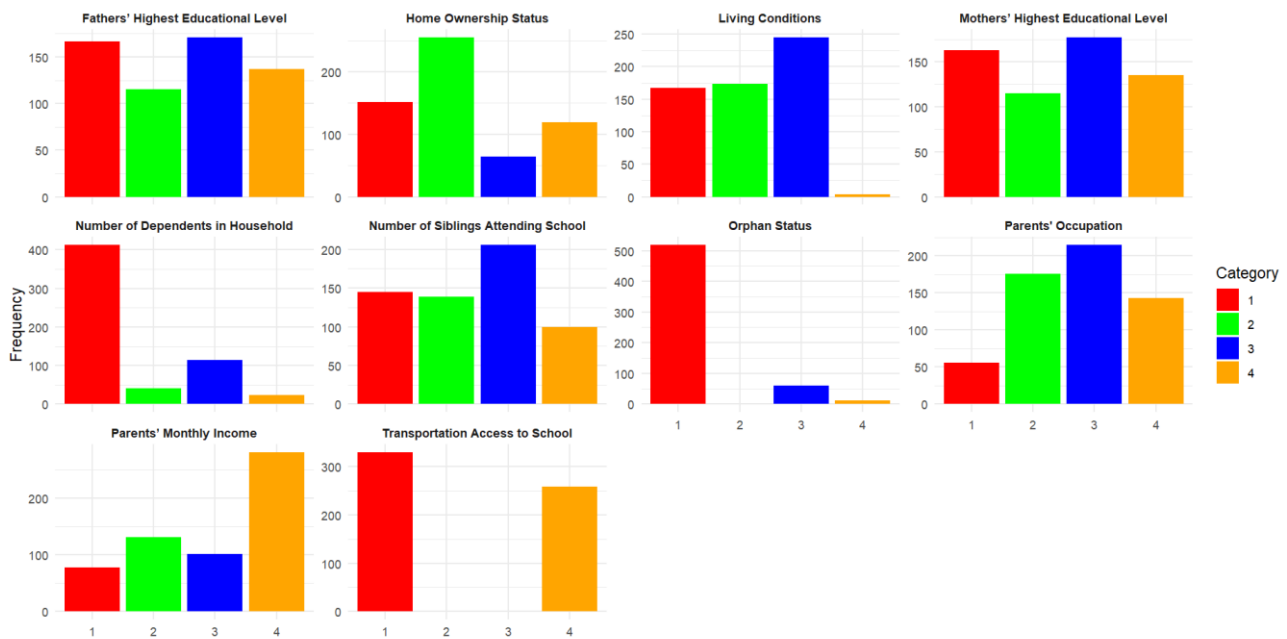


Figure 1. Bar Plot of Research Variable

4. K-Means Clustering Procedure

The clustering analysis in this study was conducted using the K-Means algorithm to group students according to similarities in their socioeconomic vulnerability profiles. The analysis was implemented in the R statistical environment, while cluster evaluation and validation were supported using the “cluster” package. The input for the clustering process consisted of the final student-level dataset represented by ten scored social indicators, where each observation corresponded to one student and each variable had previously been transformed into an ordinal score ranging from 1 to 4. These scored indicators formed a numerical feature matrix suitable for partition-based clustering.

K-Means clustering was selected because it is an efficient unsupervised learning algorithm for partitioning observations into relatively homogeneous groups based on their similarity across multiple variables. In the context of this study, the method was considered appropriate because the objective was not to predict predefined labels, but rather to identify latent group structures among students based on multidimensional social vulnerability indicators. By grouping students with similar socioeconomic conditions, the resulting clusters could then be interpreted as different levels of priority for the Free Nutritious Meal Program.

The clustering process began by specifying a candidate number of clusters, denoted by K , and partitioning the observations into K groups. For each value of K , the K-Means algorithm iteratively assigned each student to the nearest cluster centroid and then recalculated the centroid as the average of all observations belonging to that cluster. The procedure was repeated until the cluster memberships stabilized and no substantial changes in centroid positions occurred. In this study, similarity between observations was assessed using the Euclidean distance, which is the standard distance measure commonly used in K-Means clustering for numerical data.

The purpose of clustering is to partition observations into homogeneous groups by minimizing within-cluster variation while maximizing between-cluster separation, thereby ensuring high similarity among objects within the same cluster and low similarity between different clusters (Jaeger & Banks, 2023). The algorithm for the K-Means method is as follows.

- a. Determine K as the number of clusters formed.
- b. The following formula is used to generate K centroids or the initial cluster centre point randomly, then calculate the i -th centroid cluster.

$$c_i = \frac{\sum_{i=1}^n x_i}{n} \quad (1)$$

where, c_i is the centroid of the cluster, x_i is the i -th observation object, and n is the number of objects that are members of the cluster.

- c. Calculate the distance of each object to each centroid in each cluster with the following equation.

$$d(x_i, y_i) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (2)$$

- d. Allocate each object to the nearest centroid.
- e. Iterate again, then determine the position of the new centroid using equation (1).
- f. Repeat step 3 if the new centroid position is different.

5. Determination of the Optimal Number of Clusters Using the Silhouette Method

The determination of the optimal number of clusters is an important stage in clustering analysis because different values of K may produce different grouping structures and interpretations. In this study, the Silhouette Method was used as an internal cluster validation approach to evaluate the quality of clustering solutions generated by the K-Means algorithm. The Silhouette Method measures how well each observation fits within its assigned cluster compared with other clusters, thereby assessing both within-cluster cohesion and between-cluster separation. For each candidate value of K , the data are partitioned into K clusters, and the silhouette coefficient is calculated for every observation. The silhouette value for observation i is defined as:

$$s(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}} \quad (3)$$

where $a(i)$ represents the average distance between observation i and all other observations in the same cluster, while $b(i)$ denotes the minimum average distance between observation i and observations in the nearest neighboring cluster.

The silhouette coefficient ranges from -1 to 1 . A value close to 1 indicates that an observation is well matched to its own cluster and clearly separated from other clusters, whereas a value close to 0 indicates that the observation lies near the boundary between two clusters. Negative values suggest that an observation may have been assigned to an inappropriate cluster. To determine the optimal number of clusters, the average silhouette coefficient is computed for each candidate clustering solution, and the value of K with the highest average silhouette width is generally selected as the most appropriate clustering structure. Therefore, the Silhouette Method is widely used in clustering studies because it provides an interpretable and quantitative criterion for selecting the number of clusters while balancing compactness within clusters and separation between clusters.

C. RESULT AND DISCUSSION

1. Data Distribution Analysis

The results showed that most students involved in this study came from families with moderate to low economic backgrounds. Several schools demonstrated higher levels of social vulnerability based on the collected indicators, as shown in Table 2.

Table 2. Distribution of Student Participants

No	School Category	Number of Students
1	Senior High School	320
2	Vocational High School	273
Total		593

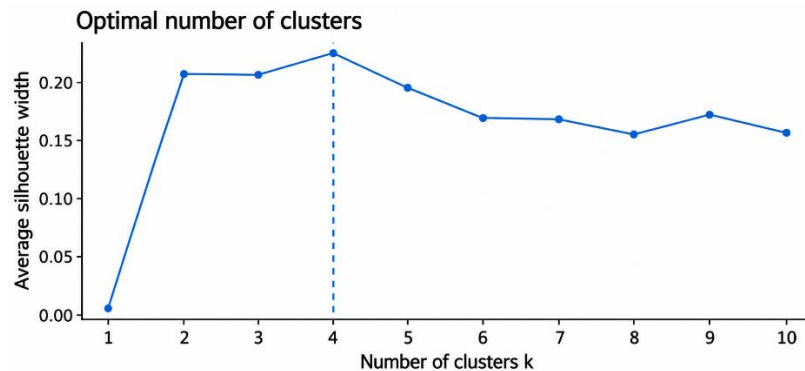
The dataset represents a diverse population of students with varying socioeconomic characteristics. The collected data provide a comprehensive basis for analyzing vulnerability patterns and identifying schools requiring priority intervention through the Free Nutritious Meal Program.

2. Comparison of Clustering Methods

Several clustering methods were considered in this study, including Hierarchical Clustering, DBSCAN, Fuzzy C-Means, and K-Means Clustering. The comparison was conducted based on computational efficiency, interpretability, scalability, and suitability for multidimensional socioeconomic data. Based on these considerations, K-Means Clustering was selected because it provides efficient computation, simple implementation, and clear cluster interpretation for educational social vulnerability classification. The method is also highly suitable for datasets with numerical scoring variables generated from social indicators. The clustering process used the K-Means Clustering method with optimization of the number of clusters using the Silhouette Method. K-Means clustering is widely recognized as an efficient unsupervised learning algorithm for grouping data with similar characteristics (Rokach & Maimon, 2020). Meanwhile, the Silhouette Method is commonly used to evaluate clustering quality and determine the optimal number of clusters (Sharma & Sharma, 2021). The silhouette score value of 0.234 indicated that the optimal number of clusters was four, as shown in Table 3.

Table 3. Comparison of Clustering Methods

Method	Advantages	Limitations
Hierarchical Clustering	Produces hierarchical grouping structure	High computational complexity
DBSCAN	Detects noise and outliers effectively	Sensitive to parameter selection
Fuzzy C-Means	Allows flexible cluster membership	Complex interpretation
K-Means Clustering	Fast, simple, and efficient for large numerical datasets	Sensitive to centroid initialization

**Figure 2.** Optimal Cluster Determination Using the Silhouette Method

Because K-Means is sensitive to the selection of initial centroids, the algorithm was run iteratively for several candidate values of K . In this study, we applied the candidate value with range 1-10 and the resulting partitions were evaluated using the Silhouette method to identify the most appropriate clustering structure. The final cluster solution was selected based on the highest average silhouette coefficient, indicating the best balance between within-cluster cohesion and between-cluster separation. The clustering output was subsequently interpreted by comparing the average indicator scores within each cluster to identify groups of students with high, moderate, and low levels of socioeconomic vulnerability. The silhouette coefficient is one of the most widely used internal cluster validation metrics because it measures the cohesion of observations within clusters and their separation from neighbouring clusters (Liu et al., 2022). According to figure 2, the optimal cluster solution was selected based on the highest average silhouette score. The analysis indicated that four clusters provided the most appropriate partitioning structure with a silhouette coefficient of 0.234.

3. Cluster Characteristics

The cluster characteristics are presented in Table 4.

Table 4. Cluster Characteristics

Social Indicators	Cluster 1	Cluster 2	Cluster 3	Cluster 4
Parents' Monthly Income	3.81	2.08	2.36	3.80
Parents' Occupation	3.44	1.96	2.23	3.45
Father's Education	3.41	1.49	1.73	3.31
Mother's Education	3.39	1.55	1.75	3.28
Home Ownership Status	2.70	1.71	1.71	2.90
Number of Dependents	3.12	1.61	2.14	2.87
Transportation Access	4.00	1.01	4.00	1.00
KIP/KKS/PKH Ownership	3.76	1.22	2.11	3.58

Social Indicators	Cluster 1	Cluster 2	Cluster 3	Cluster 4
Orphan Status	1.32	1.15	1.34	1.27
Living Conditions	2.45	1.80	2.12	2.30
Number of Siblings Attending School	2.51	2.46	2.38	2.40

Table 4 presents the centroid values of each socioeconomic indicator across the four identified clusters. The centroid analysis reveals distinct socioeconomic profiles that characterize different levels of educational vulnerability. Cluster 1 exhibits the highest average scores for parental income, parental occupation, educational attainment, transportation accessibility, and social assistance ownership. These characteristics indicate that students in this cluster predominantly originate from economically disadvantaged households with limited educational resources and restricted access to transportation. Consequently, Cluster 1 represents the highest socioeconomic vulnerability group and is categorized as Highly Priority for the implementation of the Free Nutritious Meal Program. In contrast, Cluster 4 demonstrates slightly lower vulnerability levels than Cluster 1 while still maintaining relatively high scores across most socioeconomic indicators. Students within this cluster generally experience multiple socioeconomic disadvantages, although the severity is comparatively lower. Therefore, Cluster 4 is categorized as the Priority group. Cluster 3 is characterized by moderate values across most indicators, suggesting intermediate socioeconomic conditions. Families within this cluster generally experience partial economic constraints without exhibiting the extreme deprivation observed in Cluster 1. Accordingly, Cluster 3 represents the Moderately Priority category. Meanwhile, Cluster 2 consistently records the lowest average scores for nearly all indicators, reflecting relatively stable household economic conditions, better parental educational attainment, improved transportation accessibility, and lower dependence on social assistance programs. Therefore, this cluster is classified as the Non-Priority group.

The centroid patterns demonstrate that parental economic conditions remain the dominant determinant of school priority. Nevertheless, the clustering results also reveal that educational vulnerability cannot be adequately explained by household income alone. Transportation accessibility, parental educational attainment, housing conditions, and participation in government social assistance programs collectively contribute to the formation of vulnerability profiles. This finding supports the concept of multidimensional social vulnerability, which emphasizes that educational disadvantage emerges from the interaction of multiple socioeconomic factors rather than a single economic indicator.

4. Priority School Categories

The Priority School Categories are presented in Table 5.

Table 5. Priority School Categories

School Code	Highly Priority	Priority	Moderately Priority	Non-Priority
School A	0%	2%	31%	67%
School B	63%	24%	8%	5%
School C	0%	4%	19%	77%
School D	14%	15%	42%	28%
School E	0%	3%	33%	64%

School Code	Highly Priority	Priority	Moderately Priority	Non-Priority
School F	8%	14%	32%	46%
School G	41%	53%	2%	5%
School H	51%	49%	0%	0%

The clustering results demonstrated variations in school priorities. School B and School H were categorized as highly prioritized schools because the majority of their students belonged to the Highly Priority cluster (Cluster 1). Meanwhile, School G was categorized as a priority school because most of its students belonged to the Priority cluster, whereas School D was classified as a moderately prioritized school due to the predominance of students in the Moderately Priority cluster. Schools A, C, E, and F were categorized as non-priority schools because most of their students were grouped into the non-priority cluster. Similar findings were reported Alqahtani et al. (2022), who demonstrated that clustering techniques can effectively identify vulnerable educational groups for social assistance allocation. Furthermore, Rahman et al. (2020) emphasized that clustering algorithms are capable of improving classification accuracy in educational decision-support systems. These findings are consistent with previous studies showing that clustering techniques are effective for social assistance classification and educational decision support systems (Nasution et al., 2024; Putra & Kurniawan, 2023).

The proposed framework should not be interpreted as replacing the existing government prioritization policy. Instead, it complements current criteria by providing an objective quantitative assessment of students' socioeconomic vulnerability. By integrating multiple socioeconomic indicators into a clustering model, policymakers can obtain additional evidence to support more consistent and transparent prioritization decisions, particularly in situations where schools exhibit similar regional characteristics but differ in the socioeconomic conditions of their students.

D. CONCLUSION AND SUGGESTIONS

This study successfully implemented the K-Means Clustering algorithm to identify priority schools for the Free Nutritious Meal Program in Pinrang Regency based on multidimensional socioeconomic vulnerability indicators. The clustering analysis demonstrated that integrating multiple socioeconomic variables—including household economic conditions, parental educational background, housing characteristics, transportation accessibility, and social assistance status—provides a more comprehensive basis for prioritizing schools than relying on a single socioeconomic indicator. The Silhouette Method identified four clusters as the optimal clustering solution, enabling schools to be categorized according to different levels of socioeconomic vulnerability.

The proposed clustering framework generated meaningful socioeconomic groupings and provided an objective, transparent, and data-driven basis for identifying schools with relatively higher concentrations of vulnerable students. Consequently, the proposed approach can serve as a decision-support tool for policymakers in improving the effectiveness, fairness, and transparency of the Free Nutritious Meal Program and other educational welfare initiatives. Future research is recommended to compare the performance of K-Means with other clustering algorithms, such as Hierarchical Clustering, Fuzzy C-Means, DBSCAN, and Gaussian Mixture

Models, to determine the most suitable approach for socioeconomic vulnerability assessment. Furthermore, hybrid clustering approaches that combine multiple algorithms or optimization techniques should be explored to improve clustering quality and cluster separation. Future studies should also employ larger and more diverse datasets from different regions to evaluate the robustness, generalizability, and consistency of the proposed framework across various socioeconomic and geographical contexts.

ACKNOWLEDGEMENT

The authors would like to express their sincere gratitude to the Kementerian Pendidikan Tinggi, Sains, dan Teknologi Republik Indonesia for funding this research through the Beginner Lecturer Research (PDP) grant scheme. under Contract Number: 014/C3/DT.05.00/PL/2025 The authors also extend their appreciation to LPPM-PM Institut Teknologi Bacharuddin Jusuf Habibie for the administrative support and research facilitation provided during the implementation of this study. In addition, the authors would like to thank all participating schools and respondents for their valuable contributions to this research

REFERENCES

- Al Aldowah, H., Al-Samarraie, H., & Fauzy, W. M. (2019). Educational data mining and learning analytics for 21st century higher education: A review and synthesis. *Telematics and Informatics*, 37, 13–49. <https://doi.org/10.1016/j.tele.2019.01.007>
- Ajtai, I., Ștefănie, H., Costache, R., Mihai, B., Bui, D. T., & Baciuc, C. (2023). *Mapping social vulnerability to floods: A comprehensive framework using a vulnerability index approach and PCA analysis*. Ecological Indicators.
- Alqahtani, A., Rajkhan, A., & Aligery, A. (2022). Social vulnerability analysis using clustering techniques for educational assistance programs. *IEEE Access*, 10, 45621–45633. <https://doi.org/10.1109/ACCESS.2022.3167428>
- Atif, M., Shafiq, M., Farooq, M., Ayub, G., Leisch, F., & Ilyas, M. (2023). Monitoring changes in clustering solutions: A review of models and applications. *Journal of Probability and Statistics*, 2023, 7493623. <https://doi.org/10.1155/2023/7493623>
- Barnabas, B., Bavorova, M., & Madaki, M. Y. (2024). The effect of school feeding program on pupils' enrolment, attendance and performance in Northeastern Region of Nigeria. *SAGE Open*, 14(4). <https://doi.org/10.1177/21582440241255809>
- Bittencourt, J. C. N., Costa, D. G., Portugal, P., & Vasques, F. (2024). A data-driven clustering approach for assessing spatiotemporal vulnerability to urban emergencies. *Sustainable Cities and Society*, 108, 105477. <https://doi.org/10.1016/j.scs.2024.105477>
- Chaudhry, M., Shafi, I., Mahnoor, M., Vargas, D. L. R., Thompson, E. B., & Ashraf, I. (2023). A systematic literature review on identifying patterns using unsupervised clustering algorithms: A data mining perspective. *Symmetry*, 15(9), 1679. <https://doi.org/10.3390/sym15091679>
- Fitzgerald, A., Avirmed, T., & Battulga, N. (2024). Exploring the factors informing educational inequality in higher education: A systematic literature review. *International Journal of Inclusive Education*. <https://doi.org/10.1080/13603108.2024.2381121>
- Follett, L., & Henderson, H. (2023). A hybrid approach to targeting social assistance. *Journal of Development Economics*, 160, 103002. <https://doi.org/10.1016/j.jdeveco.2022.103002>
- Hossin, M. A., Du, J., Mu, L., & Asante, I. O. (2023). Big data-driven public policy decisions: Transformation toward smart governance. *SAGE Open*, 13(4). <https://doi.org/10.1177/21582440231215123>
- Kaur, A., Kumar, Y., & Sidhu, J. (2024). Exploring meta-heuristics for partitional clustering: Methods, metrics, datasets, and challenges. *Artificial Intelligence Review*, 57, 287. <https://doi.org/10.1007/s10462-024-10920-1>
- Liu, T., Yu, H., & Blair, R. H. (2022). Stability estimation for unsupervised clustering: A review. *WIREs Computational Statistics*, 14(6), e1575. <https://doi.org/10.1002/wics.1575>

- Lu, Y., Yeom, S., Maktoubian, J., Rahman, M. M., & Kim, S.-H. (2025). Improve Student Risk Prediction with Clustering Techniques: A Systematic Review in Education Data Mining. *Education Sciences*, 15(12), 1695. <https://doi.org/10.3390/educsci15121695>
- Mah, J. C., Penwarden, J. L., Pott, H., Theou, O., & Andrew, M. K. (2023). Social vulnerability indices: A scoping review. *BMC Public Health*, 23(1), 1–11. <https://doi.org/10.1186/s12889-023-16097-6>
- Majumder, S., Roy, S., Bose, A., & Roy Chowdhury, I. (2023). Multiscale GIS based-model to assess urban social vulnerability and associated risk: Evidence from 146 urban centers of Eastern India. *Sustainable Cities and Society*, 96, 104692. <https://doi.org/10.1016/j.scs.2023.104692>
- Nasution, M. I., Siregar, B., & Harahap, M. K. (2024). Clustering analysis for social assistance recipient classification in educational institutions. *Heliyon*, 10(5), e24015. <https://doi.org/10.1016/j.heliyon.2024.e24015>
- Palermos, D., Pavi, E., Halvatsiotis, P., Mangoulia, P., Sergeantanis, T. N., & Psaltopoulou, T. (2024). Individual-based socioeconomic vulnerability and deprivation indices: A scoping review. *Frontiers in Public Health*, 12, 1403723. <https://doi.org/10.3389/fpubh.2024.1403723>
- Putra, A. R., & Kurniawan, D. (2023). Data-driven educational assistance prioritization using clustering methods. *Sustainability*, 15(4), 3567. <https://doi.org/10.3390/su15043567>
- Rahman, M. M., Islam, M. A., & Hossain, M. S. (2020). K-means clustering approach for student performance analysis and prediction. *International Journal of Advanced Computer Science and Applications*, 11(6), 493–499. <https://doi.org/10.14569/IJACSA.2020.0110662>
- Rokach, L., & Maimon, O. (2020). *Data mining with decision trees: Theory and applications* (3rd ed.). World Scientific.
- Romero, C., & Ventura, S. (2020). Educational data mining and learning analytics: An updated survey. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 10(3). <https://doi.org/10.1002/widm.1355>
- Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, 20, 53–65. [https://doi.org/10.1016/0377-0427\(87\)90125-7](https://doi.org/10.1016/0377-0427(87)90125-7)
- Sharma, A., & Sharma, S. (2021). Applications of clustering techniques in educational data mining: A systematic review. *Journal of Big Data*, 8(1), 1–28. <https://doi.org/10.1186/s40537-021-00442-3>
- Singh, S., Kumar, S., & Sharma, R. (2023). Big data analytics and e-governance: Actors, opportunities, tensions, and applications. *Technological Forecasting and Social Change*, 193, 122612. <https://doi.org/10.1016/j.techfore.2023.122612>
- Tan, C. Y. (2024). Socioeconomic status and student learning: Insights from an umbrella review. *Educational Psychology Review*, 36, Article 100. <https://doi.org/10.1007/s10648-024-09929-3>
- Wang, D., Shinde, S., Young, T., & Fawzi, W. W. (2021). Impacts of school feeding on educational and health outcomes of school-age children and adolescents in low- and middle-income countries: A systematic review and meta-analysis. *Journal of Global Health*, 11, 04051. <https://doi.org/10.7189/jogh.11.04051>
- Wani, A. A. (2024). Comprehensive analysis of clustering algorithms: Exploring limitations and innovative solutions. *PeerJ Computer Science*, 10, e2286. <https://doi.org/10.7717/peerj-cs.2286>
- Zhang, L., Clay, O. J., Lee, S.-Y., & Howell, C. R. (2024). Analyzing Multiple Social Determinants of Health Using Different Clustering Methods. *International Journal of Environmental Research and Public Health*, 21(2), 145. <https://doi.org/10.3390/ijerph21020145>