Vol. 8, No. 2, April 2026

ISI Journal of Information Systems and Informatics

Published By Asosiasi Doktor Sistem Informasi Indonesia

Data Mining Analysis for KIP Scholarship Eligibility Using Integrated DBSCAN and TOPSIS

Imam Akbar[1], **Chyquitha Danuputri**[2], **Rahma**[3], **Ita Sarmita Samad**[4]

[1]Universitas Muhammadiyah Enekang, Jl. Jenderal Sudirman No. 17 Enrekang
[2]Universitas Muhammadiyah Makassar, Jl. Sultan Alauddin No. 259, Gunung Sari, Rappocini, Makassar
[3,4]State University of Makassar, Jl. Dg. Tata Raya, Kampus UNM Parangtambung, Makassar

Received:
November 23, 2025
Revised:
March 10, 2026
Accepted:
March 27, 2026
Published:
April 12, 2026

Corresponding Author:
Author Name*:
Imam Akbar
Email*:
imamakbar071093@gmail.com

DOI:
10.63158/journalisi.v8i2.1534

Abstract. This study aims to objectively analyze the feasibility of prospective recipients of the Smart Indonesia Card Scholarship (KIP-K) by integrating the Density-Based Spatial Clustering of Applications with Noise (DBSCAN) algorithm and the Technique for Order Preference by Similarity to Ideal Solution (TOPSIS) method. The research dataset consists of 287 data on prospective scholarship recipients with 11 main attributes that reflect the socio-economic and academic conditions of students. The research process includes data collection, pre-processing, transformation of categorical attributes into numerical values using a linear weighting scheme, cluster analysis using DBSCAN, and candidate ranking using TOPSIS. DBSCAN is used to identify cluster patterns and detect anomalies in the data of potential recipients, while TOPSIS is used to rank candidates based on proximity to the ideal solution. The results of the grouping produced 10 clusters and one noise cluster that showed a variety of socio-economic characteristics of prospective scholarship recipients. The results of the ranking show that some of the candidates with the highest TOPSIS scores come from clusters with higher levels of economic vulnerability. In addition, some of the high-scoring candidates also came from the noise cluster, indicating that even though they did not belong to a particular group, they still met the eligibility criteria based on a multi-criteria evaluation. These findings show that the combination of DBSCAN and TOPSIS has the potential to support the process of analyzing the eligibility of scholarship recipients in a more systematic and data-driven manner.

Keywords: Data mining; DBSCAN; KIP Scholarship; TOPSIS; Multi-Criteria analysis

1. INTRODUCTION

Higher education has a strategic role in supporting the achievement of the Sustainable Development Goals (SDGs), especially in efforts to reduce poverty and improve the quality of education. Wider access to education is expected to improve the quality of human resources and create fairer opportunities for people from various socio-economic backgrounds[1], [2], [3]. Therefore, various countries, including Indonesia, have developed education policies aimed at expanding access to higher education for underprivileged groups. One of the policies implemented by the Indonesian government is the Indonesia Smart College Card (KIP-K) program which aims to provide tuition assistance to students from low-income families but with good academic potential[4], [5], [6].

This program is expected to reduce the gap in access to higher education and help students complete their studies on time [7], [8]. Through this program, the government seeks to ensure that economic limitations are not an obstacle for prospective students to obtain higher education. Universitas Muhammadiyah Enrekang (UNIMEN) as one of the universities in South Sulawesi also supports the implementation of the KIP-K program as part of efforts to improve the quality of human resources in the region. Data from the UNIMEN New Student Admissions Team shows that every year more than 90% of prospective new students create a KIP-K account as an effort to obtain tuition assistance. This shows that most prospective students come from economic backgrounds who need financial support to pursue higher education [10].

However, the number of KIP-K scholarship recipients allocated to UNIMEN each year is relatively limited compared to the number of applicants. In 2022, the available recipient quota is only 22 students, then decreased to 13 students in 2023 and 15 students in 2024. On the other hand, the number of applicants reaches more than 300 prospective students every year. The imbalance between the number of applicants and the available quota causes the selection process of scholarship recipients to be a challenge for universities in determining the candidates who are most deserving of assistance. In addition to quota limitations, the selection process for scholarship recipients in many universities is still carried out conventionally and has not fully utilized a data-based analysis approach. Manual selection procedures have the potential to cause subjectivity, inconsistencies in assessment, and limitations in comprehensively processing various socio-economic criteria [11][12][13]. This condition shows the need for a more systematic and data-based approach to support a more objective, transparent, and accountable scholarship recipient selection process. Along with the development of information technology, various data mining and multi-criteria decision making (MCDM) approaches have been widely used to support the decision-making process in various fields, including in the selection of scholarship recipients. One of the algorithms that can be used to analyze data patterns is Density-Based Spatial Clustering of Applications with Noise (DBSCAN). This algorithm has the ability to group data by density and detect anomalies or outliers without requiring a predetermined number of clusters [14].

In the context of scholarship selection, this ability is important to identify the characteristic patterns of prospective recipients as well as detect candidates who have different conditions from the majority of the data. In addition, the Technique for Order Preference by Similarity to Ideal Solution (TOPSIS) method is one of the multicriteria decision-making techniques that is widely used to rank alternatives based on their proximity to positive and negative ideal solutions. This method allows the integration of various criteria that are both profit and cost, such as academic achievement, parental income, number of family dependents, and asset ownership [15]. By considering various criteria simultaneously, this method can help produce more objective decisions in determining the priorities of scholarship recipients.

Several previous studies have applied various methods to determine scholarship recipients, such as Simple Additive Weighting (SAW), Weighted Product (WP), Analytical Hierarchy Process (AHP), as well as clustering techniques such as K-Means and K-Medoids [16][17], [18][19]. These methods are generally used to rank or group data separately. However, most of the research still focuses on one analytical approach only, so it is not yet fully able to identify data patterns as well as conduct a comprehensive evaluation of candidate feasibility. In addition, research that integrates density-based clustering techniques with multi-criteria decision making methods in the context of scholarship recipient selection is still relatively limited. In particular, there have not been many studies that combine the DBSCAN algorithm with the TOPSIS method to analyze the eligibility of KIP-K scholarship recipients. In fact, the combination of these two approaches has the potential to provide a more comprehensive analysis by identifying the characteristic patterns of prospective scholarship recipients while objectively ranking candidates based on various socio-economic and academic criteria. Based on these problems, this study aims to analyze the feasibility of prospective KIP-K scholarship recipients by integrating the DBSCAN algorithm and the TOPSIS method. The integration of these two methods is expected to provide a more systematic analytical approach in identifying the characteristic patterns of prospective scholarship recipients and determining the priorities of recipients based on multi-criteria evaluation. The results of this research are expected to support more objective, transparent, and data-based decision-making in the selection process of scholarship recipients in higher education.

2. METHODS

2.1. Research Design

This study uses a data mining and multi-criteria decision making (MCDM) approach to analyze the eligibility of prospective recipients of the Indonesia Smart College Card (KIP-K) scholarship. The method used is an integration between the Density-Based Spatial Clustering of Applications with Noise (DBSCAN) algorithm [20]for data pattern analysis and the Technique for Order Preference by Similarity to Ideal Solution (TOPSIS) [21] for candidate ranking. In general, the research process consists of several main stages as shown in Figure 1, which describes the stages of analysis ranging from data acquisition to interpretation of results.

2.2. Data Collection

The dataset used in this study was obtained from the KIP-K Management Information System and the New Student Admission (PMB) database of the University of Muhammadiyah Enrekang. The dataset consists of 287 data on prospective scholarship recipients with several attributes that represent the socio-economic and academic conditions of students. The variables used in this study can be seen in the Table 1. These variables in Table 1 were chosen because they represent indicators of socioeconomic vulnerability and academic ability of prospective scholarship recipients.

Table 1. Data Attributes

No	Variabel
1	Status DTKS

No	Variabel
2	Status P3KE
3	Parenting Work
4	Parents' income
5	Parental status
6	Number of family dependents
7	Home ownership
8	Access to clean water
9	Sanitation facilities
10	Distance of residence
11	Academic achievement

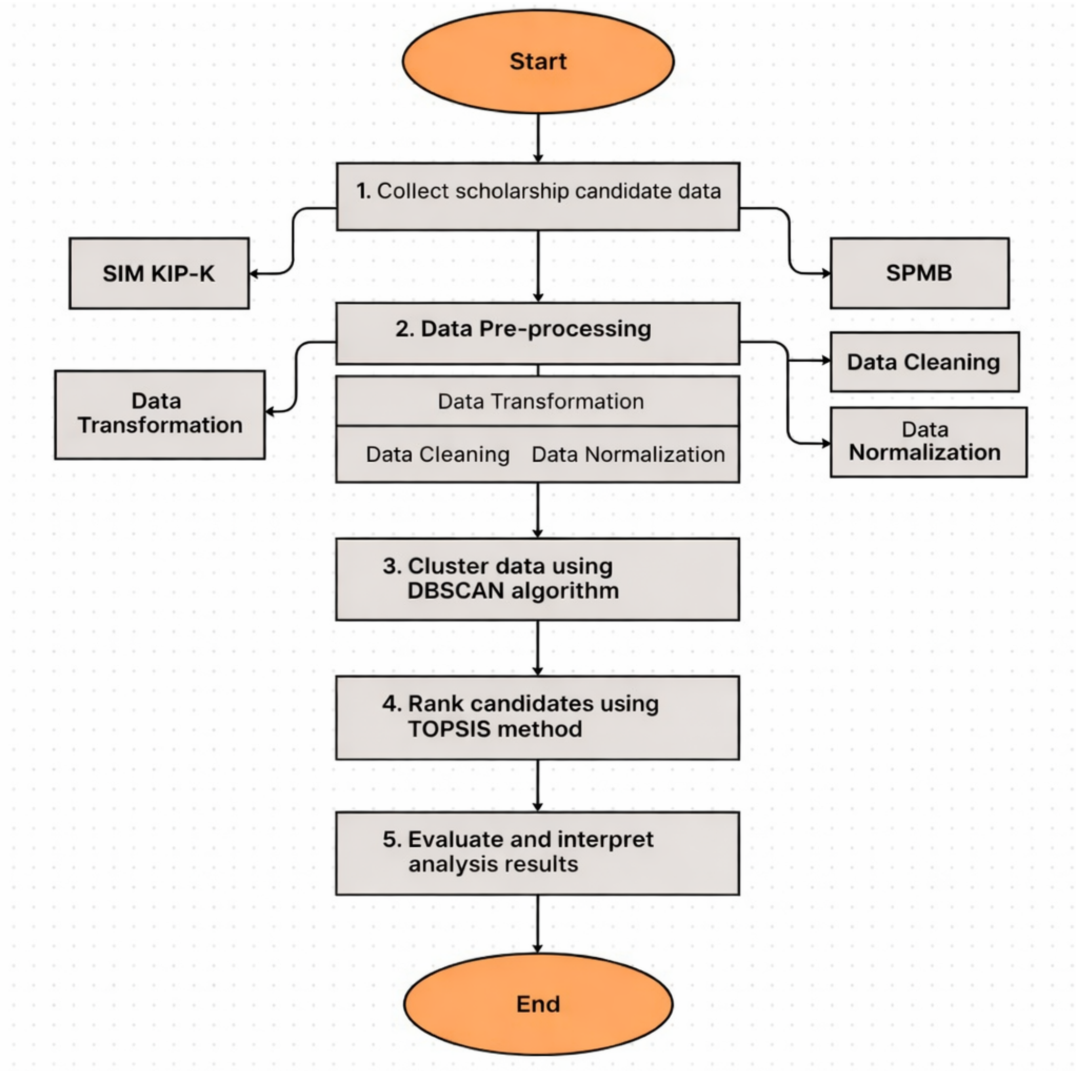


Figure 1. Research Flow Diagram

2.3. Data Preprocessing

The pre-processing stage is conducted to improve data quality before the clustering and ranking analyses are performed. This stage is essential because raw data often contain problems that can reduce the accuracy and reliability of the analysis results. The pre-processing activities include data cleaning, data transformation, and data normalization. Data cleaning is used to eliminate duplicate records and correct inconsistent values in the dataset, thereby ensuring that the data are more accurate and suitable for further analysis [22]. After the cleaning process, data transformation is carried out to convert categorical attributes into numerical values using a linear weighting scheme. This transformation is necessary because the clustering and ranking methods require data in numerical form so that each attribute can be processed and compared systematically.

The next step is data normalization, which is performed to prevent certain attributes from dominating the analysis due to differences in scale or value range. In this study, the normalization method applied is min-max normalization [23], [24], which is expressed as shown in Equation 1.

$$x' = \frac{x - x_{min}}{x_{max} - x_{min}} \tag{1}$$

where $x$represents the original value, $x_{\min}$is the minimum value of the attribute, and $x_{\max}$is the maximum value of the attribute. Through this method, all data are transformed into a standardized range between 0 and 1, allowing each attribute to contribute proportionally to the analysis. As a result, the normalized data become more balanced and appropriate for use in subsequent clustering and ranking processes.

2.4. DBSCAN Clustering

The DBSCAN clustering method, or Density-Based Spatial Clustering of Applications with Noise, was applied to identify cluster patterns in the data of prospective scholarship recipients. DBSCAN is a density-based algorithm that groups data points according to the density of their distribution in a multidimensional space, making it suitable for detecting natural groupings as well as outliers in the dataset. In this study, the algorithm uses two main parameters, namely Epsilon (ε) and Minimum Points (MinPts). The εparameter determines the neighborhood radius around a data point, while MinPts specifies the minimum number of neighboring points required for a point to be considered part of a dense region. To calculate the distance between data points, this study employs the Euclidean Distance method [25], as shown in Equation 2.

$$d(p,q) = \sqrt{\sum_{i=1}^{n} (p_i - q_i)^2} \tag{2}$$

In Equation (2), $p$denotes the first data point, $q$denotes the second data point, and $n$represents the number of attributes. The Euclidean distance value is used to determine the closeness between points and to assess whether a point lies within the neighborhood radius defined by ε.

Based on the density criteria, DBSCAN classifies data into three categories, namely core points, border points, and noise points. A core point is a point that has a number of neighboring points greater than or equal to MinPts within the radius ε. A border point is a point located within the εradius of a core point but does not have enough neighbors to qualify as a core point itself. Meanwhile, a noise point is a point that does not belong to any cluster because it does not satisfy the density requirement and is not reachable from any core point. In this study, the parameter values used are $\varepsilon = 0.6$and MinPts = 3, which were determined through parameter exploration and data distribution analysis. The clustering results were then used to identify patterns in the socioeconomic characteristics of prospective scholarship recipients, thereby providing an overview of candidate groupings based on data similarity.

After the clustering process, the TOPSIS method was used to determine the priority ranking of scholarship candidates based on their closeness to the ideal solution. TOPSIS stands for Technique for Order Preference by Similarity to Ideal Solution, and this method is widely used in multi-criteria decision-making problems. The fundamental concept of TOPSIS is that the best alternative should have the shortest distance from the positive ideal solution and the greatest distance from the negative ideal solution. The first step in TOPSIS is constructing the decision matrix, as presented in Equation 3.

$$X = [x_{ij}] \tag{3}$$

In Equation (3), x_{ij} represents the value of alternative i on criterion j. This matrix contains the performance values of each scholarship candidate for all criteria used in the evaluation process. After the decision matrix is formed, the next step is to normalize the matrix so that all criteria can be compared on the same scale. The normalization process is shown in Equation 4.

$$r_{ij} = \frac{x_{ij}}{\sqrt{\sum_{i=1}^{m} x_{ij}^2}} \tag{4}$$

As shown in Equation (4), the normalization process transforms the original values into comparable values by dividing each criterion value by the square root of the sum of squares of all alternative values in that criterion. This step is important to prevent differences in measurement scales from affecting the ranking results.

After obtaining the normalized decision matrix, the next step is to calculate the weighted normalized matrix by multiplying each normalized value by its corresponding criterion weight. This process is presented in Equation 5.

$$v_{ij} = w_j r_{ij} \tag{5}$$

In Equation (5), w_j denotes the weight of criterion j, which reflects the level of importance of each criterion in the scholarship selection process. By applying these weights, the method ensures that criteria with greater significance contribute more strongly to the final decision. Once the weighted normalized matrix is obtained, the positive ideal solution and negative ideal solution are determined. The positive ideal solution is formulated in Equation 6.

$$A^+ = (v_1^+, v_2^+, \dots, v_n^+) \tag{6}$$

while the negative ideal solution is formulated in Equation 7.

$$A^- = (v_1^-, v_2^-, \dots, v_n^-) \tag{7}$$

In Equation (6), the positive ideal solution consists of the best value for each criterion, whereas in Equation (7), the negative ideal solution consists of the worst value for each criterion. These two reference points are used to assess how close each candidate is to the ideal condition.

The next stage is to calculate the distance of each alternative from the positive and negative ideal solutions. The distance from the positive ideal solution is calculated using Equation 8.

$$D_i^+ = \sqrt{\sum(v_{ij} - v_j^+)^2} \tag{8}$$

and the distance from the negative ideal solution is calculated using Equation 9.

$$D_i^- = \sqrt{\sum(v_{ij} - v_j^-)^2} \tag{9}$$

As indicated in Equations (8) and (9), these calculations measure how far each candidate is from the best possible condition and the worst possible condition. A candidate with a smaller distance to the positive ideal solution and a larger distance to the negative ideal solution is considered more preferable. The final step in TOPSIS is to calculate the preference value, which is used as the basis for ranking candidates. This calculation is shown in Equation 10.

$$C_i = \frac{D_i^-}{D_i^+ + D_i^-} \tag{10}$$

In Equation (10), $C_i$represents the preference value of alternative i. The value of $C_i$is used to rank scholarship candidates, where a higher $C_i$value indicates that the candidate is closer to the positive ideal solution and farther from the negative ideal solution. Therefore, candidates with the highest preference values are given higher priority in the scholarship selection process. Through this approach, TOPSIS provides a structured, objective, and systematic method for ranking prospective scholarship recipients based on multiple evaluation criteria.

2.5. Integration of DBSCAN and TOPSIS

The integration of DBSCAN and TOPSIS was designed to combine data pattern analysis with multi-criteria decision-making in the scholarship selection process. This integration allows the model not only to identify the natural grouping patterns that exist in the data, but also to determine the ranking priority of each candidate in a more objective and systematic manner. In the first stage, DBSCAN is applied to identify cluster patterns and detect outliers in the dataset of prospective scholarship recipients. Through this process, candidates can be grouped based on similarities in their socioeconomic characteristics, while data points that do not belong to any group can be recognized as noise or outliers. This clustering stage is important because it provides an initial understanding of the distribution of candidate profiles and reveals hidden patterns that may not be easily identified through direct observation.

The results produced by DBSCAN then provide valuable information regarding the socioeconomic characteristics of each student group. In other words, the clustering output serves as a descriptive basis for understanding how prospective scholarship recipients are distributed across different levels or categories of socioeconomic conditions. These results help explain the structure of the data and support a more contextual interpretation of candidate conditions before the ranking process is carried out. After the clustering stage is completed, the TOPSIS method is subsequently used to rank the candidates based on their preference scores. By using TOPSIS, each candidate is evaluated according to multiple criteria and then ranked based on their closeness to the positive ideal solution and their distance from the negative ideal solution. Through this integration, the proposed model is able to combine the exploratory strength of DBSCAN with the decision-support capability of TOPSIS, thereby producing a more comprehensive approach to scholarship candidate selection.

2.6. Model Evaluation

The model evaluation was conducted by comparing the ranking results generated by the TOPSIS method with the data of scholarship recipients from the previous selection period. This validation process aims to measure the degree of conformity between the model output and the scholarship selection decisions that had already been established by the university. By performing this comparison, it becomes possible to assess whether the proposed model is capable of producing recommendations that are consistent with actual decision outcomes. The evaluation process was carried out by calculating the level of agreement between the ranking results and the previous scholarship recipient data. A higher level of conformity indicates that the model has better performance in representing the decision-making pattern applied by the institution. Therefore, this evaluation stage is essential for determining the effectiveness of the integrated DBSCAN–TOPSIS model and for verifying its potential to be used as a decision-support tool in future scholarship selection processes.

3. RESULTS AND DISCUSSION

This study analyzes data from prospective students applying for the Indonesia Smart College Card (KIP-K) scholarship at the University of Muhammadiyah Enrekang. The assessment is based on 11 eligibility criteria that reflect both the socio-economic condition of the family and the academic profile of the applicants. These criteria include DTKS status, P3KE status, parental occupation, parental income, parental status, number of family dependents, home ownership, main water source, sanitation facilities, distance from residence to the city center, and academic achievement. These variables were selected because they represent the main indicators commonly used to identify economic vulnerability and educational disadvantage, which are central considerations in scholarship allocation.

The main objective of this research is to evaluate the eligibility of prospective KIP-K recipients in a more objective, systematic, and data-driven manner. In practice, scholarship selection often involves large amounts of heterogeneous data, including categorical and numerical attributes, which can make manual assessment difficult and potentially inconsistent. To address this issue, this study combines two analytical approaches: Density-Based Spatial Clustering of Applications with Noise (DBSCAN) and the Technique for Order Preference by Similarity to Ideal Solution (TOPSIS). The DBSCAN algorithm is used to identify patterns and group applicants based on similarities in their socio-economic and academic profiles without requiring a predefined number of clusters. At the same time, DBSCAN is also capable of detecting outliers or noise, which is useful for identifying applicants whose data characteristics differ substantially from the general pattern of the population.

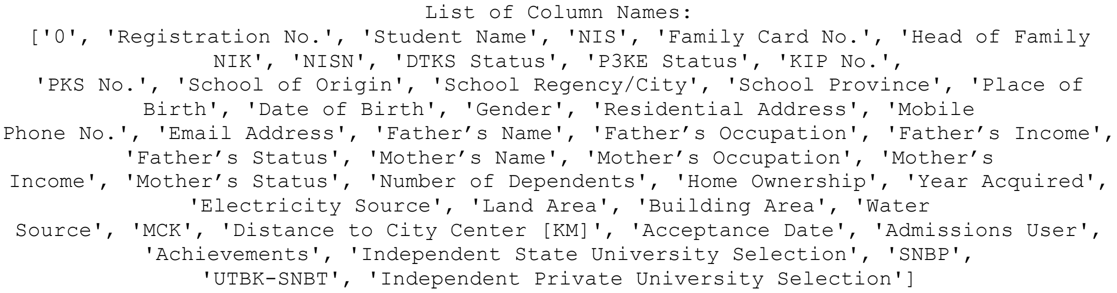
After the clustering process, the TOPSIS method is applied to rank the alternatives based on their relative closeness to the ideal solution. In this context, the ideal solution refers to the profile of applicants who most strongly meet the scholarship criteria, while the negative ideal solution represents the least eligible condition. The integration of DBSCAN and TOPSIS is particularly relevant because clustering alone does not provide a final priority order, and ranking alone may not capture hidden patterns in the dataset. Therefore, the combined use of these two methods provides a more comprehensive framework for supporting scholarship selection decisions. The results discussed in this section begin with dataset preparation, followed by data preprocessing and transformation, and finally by parameter determination for the DBSCAN algorithm.

3.1. Dataset

At the initial stage of the study, data were collected from the list of prospective students proposed as KIP-K scholarship recipients at the University of Muhammadiyah Enrekang. The research dataset was obtained from two institutional sources, namely the KIP Lecture Information System and the New Student Admission (PMB) database. These two sources provide complementary information: the KIP system contains socio-economic and scholarship-related records, while the PMB database stores applicant identity data and academic admission information. As shown in Figure 2, the dataset is organized in tabular form and contains a wide range of variables related to student background, family condition, and academic records.

Figure 2 illustrates the raw dataset structure before the analytical process was carried out. The figure shows that each row represents one applicant, while each column corresponds to a specific attribute such as income, parental occupation, family dependency, and housing characteristics. This structure is important because both DBSCAN and TOPSIS require a well-defined data matrix to perform clustering and ranking. The dataset consists of 287 applicant records with approximately 40 initial attributes, covering demographic information, household economic conditions, educational background, and academic achievement. Although not all of these attributes were used in the final model, the complete dataset provided a strong basis for selecting the most relevant variables for scholarship eligibility analysis.

The use of institutional data sources strengthens the reliability of this study because the information originates from formal admission and scholarship records rather than from unverified external sources. Data collection was conducted in a structured manner, taking into account the validity, consistency, and relevance of each variable. This is essential because scholarship recommendation systems depend heavily on the quality of input data. Inaccurate or incomplete records may lead to bias in clustering results or errors in ranking. Therefore, the dataset preparation stage plays a crucial role in ensuring that the subsequent analytical processes can produce meaningful and accountable results.



```
                          List of Column Names:
['0', 'Registration No.', 'Student Name', 'NIS', 'Family Card No.', 'Head of Family
         NIK', 'NISN', 'DTKS Status', 'P3KE Status', 'KIP No.',
'PKS No.', 'School of Origin', 'School Regency/City', 'School Province', 'Place of
      Birth', 'Date of Birth', 'Gender', 'Residential Address', 'Mobile
Phone No.', 'Email Address', 'Father’s Name', 'Father’s Occupation', 'Father’s Income',
     'Father’s Status', 'Mother’s Name', 'Mother’s Occupation', 'Mother’s
Income', 'Mother’s Status', 'Number of Dependents', 'Home Ownership', 'Year Acquired',
        'Electricity Source', 'Land Area', 'Building Area', 'Water
Source', 'MCK', 'Distance to City Center [KM]', 'Acceptance Date', 'Admissions User',
     'Achievements', 'Independent State University Selection', 'SNBP',
         'UTBK-SNBT', 'Independent Private University Selection']
```

Figure 2. The dataset obtained from the KIP-K system and PMB database

3.2. Data Preprocessing

Although the initial dataset contained around 40 attributes, only 11 key attributes were selected for further analysis. This selection was made to focus on variables that most directly represent the socio-economic vulnerability of the applicants and their academic readiness. The selected attributes are DTKS status, P3KE status, parental occupation, parental income, parental status, number of family dependents, home ownership, main water source, sanitation facilities, distance from residence to the city center, and academic achievement. These variables were considered the most relevant because they reflect both the financial hardship of the family and the educational potential of the applicant.

The preprocessing stage is essential because not all raw attributes contribute equally to the scholarship eligibility decision. Some attributes may be redundant, incomplete, or less significant in describing the core dimensions of disadvantage targeted by the KIP-K program. By reducing the dataset to the most meaningful variables, the analysis becomes more focused and computationally efficient. This process also helps minimize noise in the model, especially when applying clustering algorithms such as DBSCAN, which are sensitive to the distribution and scale of input variables. In other words, variable selection is not merely a technical step but also a conceptual effort to align the analysis with the objectives of the scholarship program.

The selected variables reflect the multidimensional nature of poverty and educational access. For example, parental occupation and parental income capture direct economic capacity, while home ownership, water source, and sanitation facilities represent living standards and access to basic services. Similarly, distance to the city center may indicate geographical disadvantage and limited access to educational resources, whereas academic achievement serves as an indicator of student merit and preparedness for higher education. By combining these dimensions, the study avoids relying on a single indicator and instead builds a more comprehensive picture of applicant eligibility.

This preprocessing stage also lays the groundwork for numerical transformation and normalization in later steps. Since several of the selected variables are categorical in nature, they cannot be directly processed by DBSCAN and TOPSIS without conversion. Therefore, after identifying the final 11 attributes, the next stage involved transforming these qualitative categories into numerical values. This transformation ensures that all selected variables can be integrated into a unified analytical framework.

3.3. Conversion of Categorical Data into Numerical Values

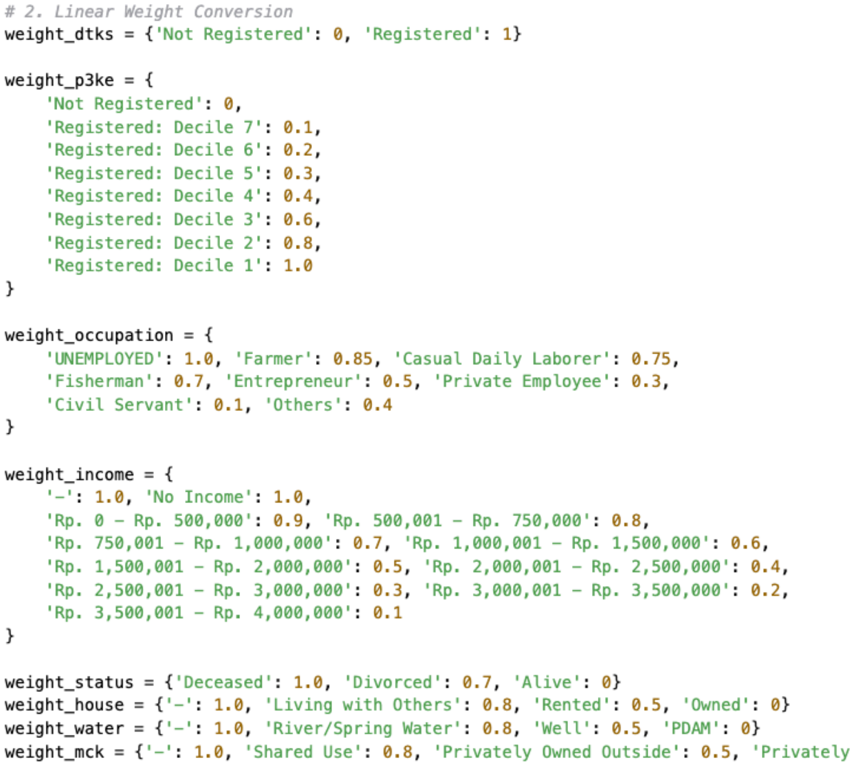
After data collection and attribute selection were completed, the next stage involved converting categorical variables into numerical values using a linear weight conversion approach. This step was necessary because many of the variables in the dataset, such as welfare status, parental occupation, and home ownership, were originally recorded as text categories. Algorithms such as DBSCAN require numerical input to calculate distances between data points, while TOPSIS also depends on numerical values to perform normalization and determine distances from the ideal and negative ideal solutions. Therefore, categorical-to-numerical transformation became a critical preprocessing step in this study.

One of the simplest examples of this transformation is the encoding of DTKS status. DTKS, or the Integrated Social Welfare Data, indicates whether a family has been formally recorded in the government welfare database. As presented in Table 2, applicants whose families are recorded in DTKS are assigned a value of 1, while those not recorded receive a value of 0. This binary conversion reflects the assumption that inclusion in DTKS is a strong indicator of economic vulnerability. The conversion framework shown in Table 2 simplifies qualitative welfare information into a format that can be used directly in computational analysis.

Table 2. DTKS Data Conversion

Category	Value
Recorded	1
Not Recorded	0

The numerical conversion process was implemented programmatically using Python. As illustrated in Figure 3, the categorical labels were mapped into corresponding numerical weights through a structured data transformation script. Referring to Figure 3 in the paragraph is important because it demonstrates that the conversion process was not conducted manually, but rather through a reproducible and systematic computational procedure. This improves methodological transparency and reduces the risk of inconsistency in assigning values across the dataset. Through this step, the study ensured that the raw categorical data could be integrated into the next analytical stages in a consistent and machine-readable format.



```python
# 2. Linear Weight Conversion
weight_dtks = {'Not Registered': 0, 'Registered': 1}

weight_p3ke = {
    'Not Registered': 0,
    'Registered: Decile 7': 0.1,
    'Registered: Decile 6': 0.2,
    'Registered: Decile 5': 0.3,
    'Registered: Decile 4': 0.4,
    'Registered: Decile 3': 0.6,
    'Registered: Decile 2': 0.8,
    'Registered: Decile 1': 1.0
}

weight_occupation = {
    'UNEMPLOYED': 1.0, 'Farmer': 0.85, 'Casual Daily Laborer': 0.75,
    'Fisherman': 0.7, 'Entrepreneur': 0.5, 'Private Employee': 0.3,
    'Civil Servant': 0.1, 'Others': 0.4
}

weight_income = {
    '–': 1.0, 'No Income': 1.0,
    'Rp. 0 – Rp. 500,000': 0.9, 'Rp. 500,001 – Rp. 750,000': 0.8,
    'Rp. 750,001 – Rp. 1,000,000': 0.7, 'Rp. 1,000,001 – Rp. 1,500,000': 0.6,
    'Rp. 1,500,001 – Rp. 2,000,000': 0.5, 'Rp. 2,000,001 – Rp. 2,500,000': 0.4,
    'Rp. 2,500,001 – Rp. 3,000,000': 0.3, 'Rp. 3,000,001 – Rp. 3,500,000': 0.2,
    'Rp. 3,500,001 – Rp. 4,000,000': 0.1
}

weight_status = {'Deceased': 1.0, 'Divorced': 0.7, 'Alive': 0}
weight_house = {'–': 1.0, 'Living with Others': 0.8, 'Rented': 0.5, 'Owned': 0}
weight_water = {'–': 1.0, 'River/Spring Water': 0.8, 'Well': 0.5, 'PDAM': 0}
weight_mck = {'–': 1.0, 'Shared Use': 0.8, 'Privately Owned Outside': 0.5, 'Privately
```

Figure 3. Python-based categorical data conversion process

3.4. P3KE Status Conversion

In addition to DTKS, the P3KE status was also transformed into numerical form. P3KE refers to the Acceleration of Extreme Poverty Eradication data, which classifies households into economic deciles. Unlike DTKS, which uses a binary status, P3KE provides a more detailed representation of economic condition by dividing households into deciles. In this study, the P3KE categories were converted using an inverse weighting approach, where applicants from lower deciles received higher numerical scores. This approach is based on the logic that lower decile positions correspond to greater economic hardship and therefore stronger eligibility for educational assistance.

As shown in Table 3, households in Decile 1 were assigned the highest score of 1.0, while households in Decile 7 received a score of 0.1. Applicants who were not registered in the P3KE database received a score of 0. By explicitly presenting this conversion in Table 3, the study provides a clear and transparent rationale for how economic ranking data were incorporated into the eligibility model. The weighting system ensures that families with the greatest socio-economic vulnerability receive proportionally higher consideration during the analytical process.

Table 3. P3KE Value Conversion

Category	Value
Decile 1	1.0
Decile 2	0.8
Decile 3	0.6
Decile 4	0.4
Decile 5	0.3
Decile 6	0.2
Decile 7	0.1
Not registered	0

The use of inverse weighting in Table 3 is consistent with the broader concept of socio-economic prioritization in educational support programs. Rather than treating all registered households equally, the method distinguishes levels of deprivation more precisely. This is especially important in scholarship selection, where limited funding often requires prioritization among many applicants who all appear eligible at a general level. By assigning higher weights to lower deciles, the model becomes more sensitive to differences in poverty intensity. As a result, the numerical representation of P3KE status contributes meaningfully to both the clustering process and the final ranking of scholarship candidates.

3.5. Transformation of Other Socio-Economic Variables

Beyond DTKS and P3KE, several additional socio-economic variables were also transformed into numerical values. These variables include parental occupation, parental income, parental status, home ownership, main water source, and sanitation facilities. The conversion of these variables followed the same general principle: categories representing greater vulnerability were assigned higher weights, while categories indicating more stable socio-economic conditions received lower weights. This weighting scheme was developed to ensure that the final numerical data reflected the real-world logic of scholarship prioritization.

For example, families in which parents have irregular employment, temporary work, or are unemployed were assigned higher values than families with stable occupations. Similarly, lower income categories received higher weights because they indicate greater financial need. Households living in rented homes or non-permanent housing conditions were also assigned higher vulnerability scores than those who owned their homes. Access to clean water and adequate sanitation was treated as an important indicator of household welfare, as inadequate facilities often reflect broader socio-economic disadvantage. By converting these categories into weighted numerical values, the study was able to incorporate qualitative household conditions into a measurable and comparable analytical framework.

This transformation was essential for two reasons. First, it allowed the construction of a multivariate numerical vector for each applicant, which is required by the DBSCAN algorithm to measure distances among data points. Second, it enabled the TOPSIS method to perform normalization and evaluate alternatives based on their relative closeness to the ideal solution. Without this transformation, many of the most meaningful indicators of socio-economic hardship could not have been included effectively in the computational model. In this sense, the transformation of socio-economic variables served as a bridge between real-world applicant conditions and quantitative decision-support analysis.

The careful assignment of linear weights also enhances interpretability. Because higher scores consistently represent greater vulnerability, the resulting dataset becomes easier to analyze and explain. This consistency is useful not only for algorithmic processing but also for institutional accountability, especially when scholarship selection decisions need to be justified. Therefore, the transformation of socio-economic variables is not only a technical necessity but also a methodological strength of the study.

3.6. Determination of DBSCAN Parameters

The performance of the DBSCAN algorithm depends strongly on the proper selection of two key parameters: eps (ε) and min_samples. The eps parameter defines the maximum distance between two data points for them to be considered neighbors, while min_samples determines the minimum number of neighboring points required to form a dense region or cluster. Choosing inappropriate values may result in either too many small clusters or an excessive number of points being classified as noise. Therefore, parameter selection is a critical step in ensuring that the clustering results accurately represent the structure of the applicant data.

To identify an appropriate value for eps, this study used a K-Distance graph, as presented in Figure 4. In the graph, the x-axis represents the ordered data points, while the y-axis shows the distance of each point to its fourth nearest neighbor (k = 4). The purpose of this graph is to identify an "elbow point," which indicates a threshold where the distance between neighboring points begins to increase sharply. Based on the visual pattern shown in Figure 4, the elbow appears in the range of approximately 0.5 to 0.6. This suggests that eps values within this interval are suitable candidates for clustering the dataset.

Referring directly to Figure 4 in the discussion is important because it provides visual justification for the parameter choice rather than relying on arbitrary selection. The graph-based approach improves methodological rigor and makes the process more transparent. Based on the pattern observed in the figure, several combinations of parameters were tested to identify the most stable and meaningful clustering configuration. The tested values included min_samples = 3, 4, 5, 6, and 7, and eps = 0.4, 0.45, 0.5, 0.55, and 0.6. By evaluating multiple parameter combinations, the study aimed to find the balance between identifying meaningful dense groups and minimizing the number of irrelevant outliers.

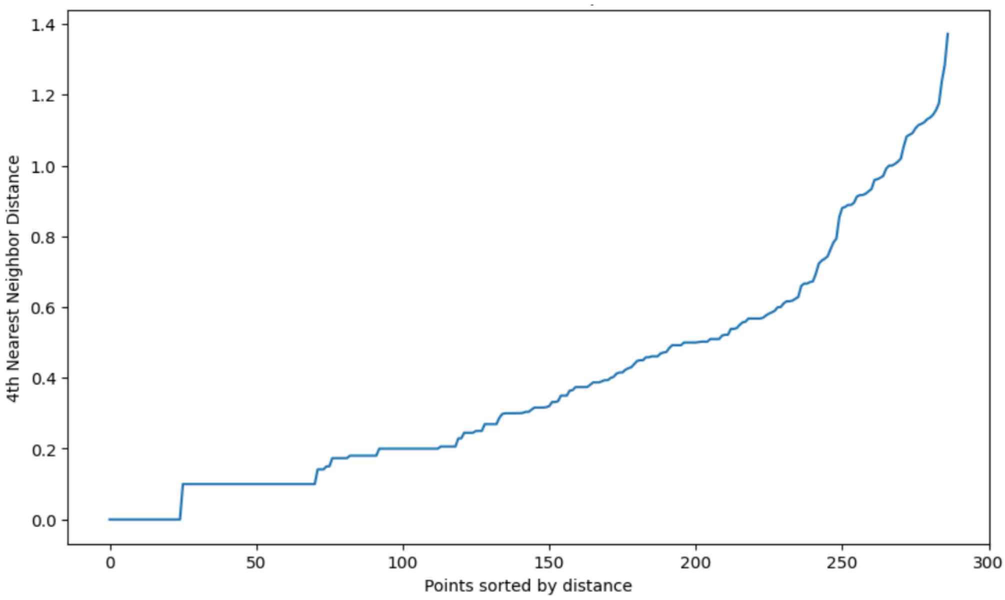


Figure 4. Determination of DBSCAN parameters using the K-Distance

This parameter exploration is especially relevant in the context of scholarship eligibility data, where applicant characteristics may not be distributed uniformly. Some groups of applicants may share very similar socio-economic profiles, while others may appear more isolated due to unique combinations of hardship indicators. DBSCAN is well suited to this type of data because it can identify naturally occurring clusters of varying density and separate unusual cases as noise. The parameter tuning process described through Figure 4 therefore forms an important foundation for the next stage of analysis, where the identified clusters can be interpreted and used to support subsequent ranking through TOPSIS.

3.7. DBSCAN Clustering

The DBSCAN clustering process was carried out using the selected optimal parameters, namely eps = 0.6 and min_samples = 3. Based on this configuration, the model produced 10 main clusters and 1 noise cluster from a total of 287 student records. The clustering result indicates that the dataset does not form a single homogeneous group, but instead consists of several natural groupings with different levels of similarity in socio-economic and academic characteristics. This is one of the main strengths of the DBSCAN method, as it is able to detect dense patterns in the data without requiring the number of clusters to be specified in advance. In addition, DBSCAN can identify observations that do not fit into any dominant pattern, which are then categorized as noise or outliers.

The visual distribution of the clustering output can be observed in Figure 5, which presents the DBSCAN results in the form of a scatter plot. In Figure 5, the horizontal axis represents the normalized distance of residence from the city center, while the vertical axis represents the father's income variable that has been converted into a linear weight. Each color in the figure corresponds to a different cluster label, including the noise cluster (-1). Referring to Figure 5, it can be seen that the data points are distributed across several cluster regions, with some groups appearing denser than others. The visualization also shows that a number of observations are scattered separately from the dominant cluster pattern, which supports their classification as outliers. This visual representation is important because it helps confirm that the grouping process was not arbitrary, but rather emerged from the intrinsic structure of the data.

The numerical distribution of cluster members is presented in Table 5. Based on Table 5, Cluster 2 is the dominant group, containing 172 students or 59.9% of the total dataset. This finding indicates that most applicants share relatively similar socio-economic profiles and are concentrated in one major cluster. Meanwhile, the noise cluster (-1) contains 34 students or 11.8%, which suggests that a notable portion of applicants have characteristics that differ from the dominant data structure. Other clusters such as Cluster 5 with 24 students (8.4%) and Cluster 0 with 23 students (8.0%) represent smaller but still meaningful subgroups. The remaining clusters contain relatively few members, showing that certain combinations of socio-economic conditions appear only in limited portions of the applicant population.

Table 5. DBSCAN Cluster

Cluster	Number of Members	Percentage
2	172	59.9%
-1 (Noise)	34	11.8%
5	24	8.4%

Cluster	Number of Members	Percentage
0	23	8.0%
4	11	3.8%
7	7	2.4%
1	4	1.4%
3	3	1.0%
6	3	1.0%
8	3	1.0%
9	3	1.0%

The dominance of Cluster 2, as shown in Table 5, suggests the presence of a large group of students with broadly comparable levels of economic vulnerability, although internal variation may still exist in variables such as parental income, housing conditions, and access to facilities. The noise cluster, on the other hand, deserves special attention because it represents applicants whose socio-economic patterns do not align closely with the majority of the dataset. In the context of scholarship selection, these outliers may correspond to applicants with either unusually severe disadvantage or unusual combinations of indicators that make them distinct from other candidates. Therefore, the DBSCAN result not only segments the applicants into meaningful groups, but also provides an additional analytical layer for understanding the heterogeneity of scholarship applicants before the final ranking stage is performed.

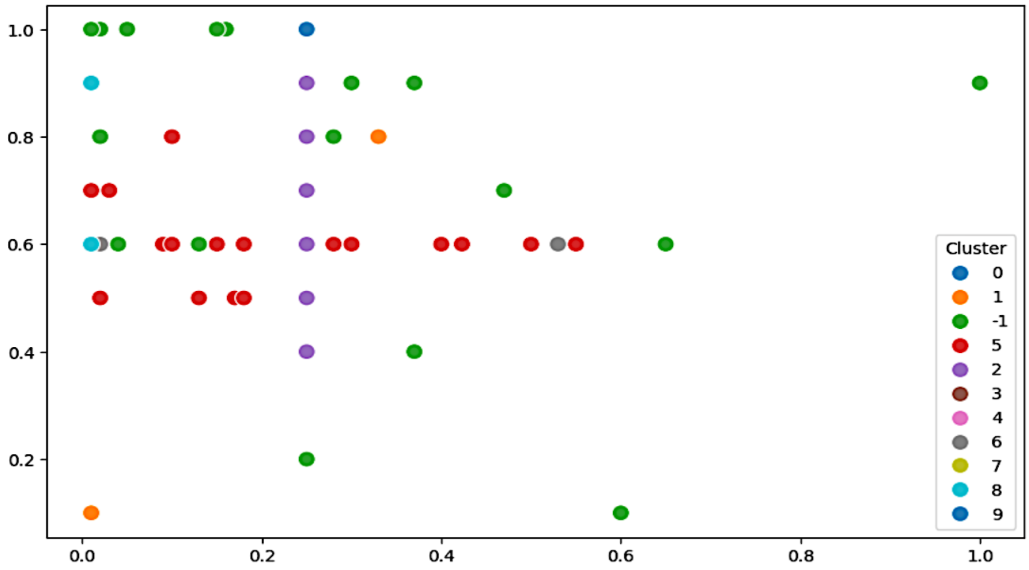


Figure 5. DBSCAN Clustering

3.8. Ranking Using TOPSIS

After the clustering stage was completed, the next step was to rank the prospective scholarship recipients using the TOPSIS method. While DBSCAN is useful for identifying natural groupings within the data, it does not provide a priority order among individual applicants. For this reason, TOPSIS was applied as a multicriteria decision-making technique to determine the relative eligibility of each student based on all selected criteria. In this method, each alternative is evaluated according to its distance from the positive ideal solution, which represents the most eligible condition, and the negative ideal solution, which represents the least eligible condition. Applicants with scores closer to the positive ideal solution are considered more feasible as scholarship recipients.

The ranking output is illustrated in Figure 6, which shows the distribution of TOPSIS scores across all alternatives. Based on the results, the TOPSIS scores range from approximately 0.30 to 0.84. Referring to Figure 6, the score pattern shows a gradual decline from the highest-ranked candidates to the lowest-ranked candidates. This indicates that the applicants are not divided into only two extreme groups, but rather spread across a continuum of eligibility levels. The visualization in Figure 6 is particularly useful because it reveals the competitive nature of the applicant pool. Instead of showing abrupt changes in every position, the graph demonstrates a relatively smooth pattern for much of the ranking, followed by sharper decline in the lower part of the list.

The upper segment of the ranking reflects the group of applicants with the strongest level of conformity to the scholarship criteria. The first 30 ranks, as indicated by the TOPSIS output shown in Figure 6, generally have scores above 0.75, which suggests that these students demonstrate a high degree of eligibility. They are likely characterized by stronger indicators of socio-economic vulnerability combined with academic profiles that support scholarship prioritization. The middle group, approximately between rank 31 and rank 150, has TOPSIS scores in the range of 0.60 to 0.70. This range indicates moderate eligibility, meaning these students still meet many of the scholarship criteria but may not show the same intensity of need or priority as the highest-ranked group. After approximately rank 260, the scores decline more sharply, which implies a lower degree of conformity with the ideal recipient profile.

The interpretation of Figure 6 confirms that TOPSIS is effective in translating multiple socio-economic and academic variables into a single composite preference score. This is especially important in scholarship selection, where decision-makers must compare many applicants across several criteria simultaneously. Rather than relying on subjective judgment alone, the TOPSIS method produces a transparent and quantifiable ranking structure. The score distribution shown in Figure 6 also helps decision-makers identify practical cut-off zones, such as highly eligible, moderately eligible, and less eligible groups. As a result, the TOPSIS ranking stage complements the clustering analysis by providing a final prioritization order that can be used directly in scholarship recommendation decisions.

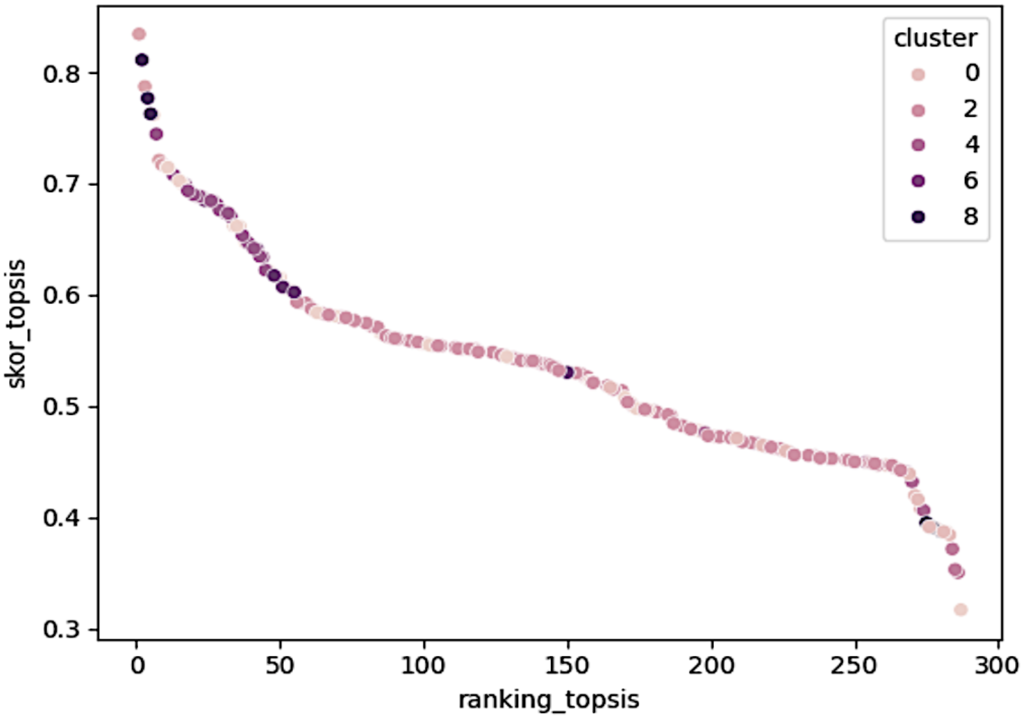


Figure 6. TOPSIS Ranking

3.9. Analysis of DBSCAN and TOPSIS Integration

The integration of DBSCAN and TOPSIS in this study provides a more comprehensive framework for evaluating KIP-K scholarship applicants. Individually, each method has a distinct analytical function. DBSCAN is designed to identify structural patterns in the dataset by grouping students with similar socio-economic characteristics and detecting outliers. TOPSIS, on the other hand, is used to rank individual applicants based on their overall closeness to the ideal scholarship recipient profile. When combined, these two methods provide both group-based insight and individual-level prioritization, which makes the selection process more systematic and informative than relying on only one approach.

The relationship between both methods can be observed from the overall analytical flow. The clustering structure generated by DBSCAN, as shown in Figure 5, reveals that the applicant population is heterogeneous and composed of several distinct socio-economic patterns. This means that scholarship applicants do not represent a uniform group, even though they are all applying under the same financial assistance scheme. Some belong to the dominant cluster, while others fall into smaller clusters or even the noise group. Once these patterns are identified, TOPSIS provides an additional layer of analysis by ranking applicants within the broader dataset according to their degree of eligibility. Thus, Figure 5 helps explain the structure of the population, while Figure 6 translates that structure into practical selection priorities.

This integration offers several important advantages. First, DBSCAN helps decision-makers understand whether certain applicants belong to dense groups of similar socio-economic conditions or whether they represent atypical cases that may require closer examination. This is particularly relevant for applicants in the noise cluster (-1) shown in Table 5, because they may have distinctive profiles that are easily overlooked in conventional manual selection. Second, TOPSIS allows the institution to assign a clear order of priority among all applicants, including those who belong to the same cluster. In other words, clustering alone can identify similarity, but it cannot determine who should be prioritized first; this limitation is addressed by TOPSIS through numerical ranking.

The combined use of these methods therefore strengthens the objectivity, consistency, and transparency of the scholarship selection process. In a manual system, decision-making may depend heavily on subjective interpretation of applicant documents, which can lead to inconsistency, especially when the number of applicants is large. By contrast, the integration of DBSCAN and TOPSIS provides a structured computational approach in which socio-economic and academic indicators are processed systematically. The cluster distribution in Table 5, the spatial grouping in Figure 5, and the preference score pattern in Figure 6 together provide a strong basis for data-driven decision-making. Therefore, this integration is not only methodologically sound, but also highly relevant for improving fairness and accountability in the determination of KIP-K scholarship recipients at the University of Muhammadiyah Enrekang.

3.10. Discussion

The results of this study indicate that the integration of the DBSCAN algorithm and the TOPSIS method provides a more comprehensive approach for evaluating prospective KIP-K scholarship recipients. This combined approach does not only produce a final ranking of candidates, but also reveals the underlying socio-economic structure of the applicant data. In conventional scholarship selection, evaluation is often carried out by checking administrative documents and comparing several basic criteria, which may not fully capture the complexity of applicants' conditions. In contrast, the use of DBSCAN allows the identification of natural groupings among students based on similarities in socio-economic and academic variables, while TOPSIS provides a systematic mechanism for prioritizing candidates within the overall population. As shown in Figure 5, the clustering process visually demonstrates that the applicant data are not distributed uniformly, but instead form several groups with different characteristics. This finding confirms that the condition of prospective scholarship recipients is multidimensional and cannot be adequately understood through a single-variable assessment.

One of the most important findings is the dominance of Cluster 2, which, according to Table 5, contains 172 students or 59.9% of the total dataset. The size of this cluster suggests that the majority of applicants share relatively similar socio-economic profiles and can be grouped into one large category of vulnerability. However, despite belonging to the same cluster, these students still show variation in several important aspects, particularly parental income, number of family dependents, and access to household facilities. This means that administrative indicators such as DTKS and P3KE status may identify broad economic vulnerability, but they do not completely explain the differences in living conditions among applicants. The result supports the view that scholarship eligibility should not rely solely on administrative registration status. Instead, it should incorporate multiple socio-economic indicators simultaneously so that the assessment reflects actual household conditions more accurately. In this regard, the concentration of applicants in Cluster 2, as presented in Table 5, highlights both the usefulness and the limitations of government welfare data when used as the sole basis for scholarship decisions.

The presence of the noise cluster (-1) also offers an important interpretation. Based on Table 5, this cluster consists of 34 students or 11.8% of the dataset, making it a substantial outlier group rather than an insignificant residual category. In DBSCAN, noise refers to observations that do not belong to any dense cluster because their characteristics differ from the dominant patterns in the data. In the context of this study, the outlier cluster likely represents applicants whose socio-economic profiles are relatively distinct from the majority, such as students with higher parental income, fewer family dependents, or less severe indicators of vulnerability. The separation of this group is also visually supported by Figure 5, where several data points appear more isolated from the denser cluster formations. This finding is meaningful for scholarship selection because it shows that DBSCAN can help identify applicants whose profiles may require additional scrutiny. In practical terms, the noise cluster may help decision-makers detect potential cases of mistargeting, administrative inconsistency, or applicants whose level of need is substantially lower than that of most candidates. Thus, the role of DBSCAN is not merely to create groups, but also to improve screening accuracy by highlighting cases that do not fit the general pattern.

Another notable result is the existence of smaller clusters, including Cluster 0, Cluster 5, and several minor clusters such as 7, 8, and 9, as reported in Table 5. These smaller clusters indicate that the applicant population includes subgroups with more specific or mixed socio-economic characteristics. For example, some students may be registered in DTKS but still display relatively better economic indicators, such as more stable parental employment or higher household income. Such a pattern may reflect an administrative anomaly, where the official welfare status has not been updated in line with the family's current condition. On the other hand, some students may not be recorded in DTKS or P3KE but still demonstrate strong indicators of economic hardship, such as limited housing ownership, inadequate sanitation, or long geographic distance from the city center. These cases reveal an important weakness in relying exclusively on administrative welfare databases. The clustering pattern shown in Figure 5 and summarized numerically in Table 5 therefore suggests that data-driven analysis can uncover hidden vulnerability that is not always visible in formal records. This has direct implications for scholarship policy, because it shows that a more flexible and analytical approach is needed to complement administrative verification.

Following the clustering stage, the application of TOPSIS adds an important decision-making layer by converting multidimensional applicant data into a ranked list of priorities. The TOPSIS results, illustrated in Figure 6, show that the scores range from approximately 0.30 to 0.84, indicating considerable variation in the degree of eligibility among applicants. The score distribution displays a relatively gradual decline from the top ranks to the middle ranks, which suggests that many candidates are competitively positioned and satisfy a substantial portion of the scholarship criteria. However, after approximately rank 260, the curve declines more sharply, indicating a lower level of conformity with the ideal recipient profile. This pattern is important because it shows that the applicant pool is not divided by a simple binary distinction between eligible and ineligible candidates. Instead, scholarship feasibility exists on a spectrum, with some students strongly matching the intended target profile and others showing only partial alignment. By referring to Figure 6, the discussion becomes more grounded because the ranking pattern can be interpreted not only numerically, but also visually as evidence of competitive dispersion among applicants.

The highest TOPSIS scores generally correspond to applicants who exhibit a combination of characteristics associated with strong scholarship eligibility, such as low family income, large numbers of dependents, limited access to basic household facilities, and supporting academic credentials. This is consistent with the objective of the KIP-K program, which is designed to provide educational opportunities to students from economically disadvantaged families while still considering student merit. What makes the TOPSIS method especially useful in this context is its ability to evaluate all relevant criteria simultaneously rather than independently. A student may be economically vulnerable but have slightly better living conditions in one dimension, while another may have severe facility limitations but lower academic readiness. TOPSIS balances these conditions by calculating each student's proximity to the positive ideal solution and distance from the negative ideal solution. Therefore, the ranking presented in Figure 6 represents a synthesized assessment rather than a one-dimensional judgment. This strengthens the fairness of the selection process because it reduces the tendency to overemphasize any single criterion.

A particularly meaningful finding in this study is that the clusters generated by DBSCAN are not always perfectly aligned with the ranking produced by TOPSIS. Some candidates with relatively high TOPSIS scores may appear in different clusters, while students within the same cluster may not share the same ranking position. This result is not a weakness of the model; rather, it reflects the fact that socio-economic vulnerability is complex, layered, and multidimensional. DBSCAN groups students based on similarity in overall data structure, while TOPSIS evaluates the degree to which each individual matches the ideal scholarship profile. In other words, clustering answers the question of who is similar to whom, whereas ranking answers who should be prioritized first. The difference between the two outputs demonstrates why a single analytical method would be insufficient for a problem as sensitive and nuanced as scholarship selection. By examining Figure 5 together with Figure 6, it becomes clear that group membership and priority ranking provide complementary, not competing, perspectives on applicant evaluation.

From a methodological standpoint, the integration of DBSCAN and TOPSIS offers a robust framework for decision support in educational assistance programs. DBSCAN contributes exploratory insight by identifying hidden patterns, dominant groups, and unusual observations in the data, while TOPSIS contributes evaluative precision by producing a clear and rational order of priority. The cluster composition presented in Table 5, the visual cluster separation in Figure 5, and the ranking trend shown in Figure 6 together create a layered analytical system that is more informative than manual review alone. This integrated model improves transparency because every stage of the process can be traced back to observable data patterns and explicit numerical calculations. It also improves consistency, since all applicants are assessed using the same criteria and computational procedures. In institutional settings where scholarship decisions may affect many students and involve limited funding quotas, this kind of systematic approach is especially valuable because it can reduce subjective bias and strengthen accountability.

In practical terms, the findings of this study have important implications for scholarship management at the University of Muhammadiyah Enrekang. The proposed model can function as a decision support system that helps the university evaluate applicants in a more structured and evidence-based manner. It does not replace human judgment entirely, but it provides a stronger analytical foundation for decision-makers when reviewing cases with similar profiles or identifying applicants who may require special consideration. The system is also useful for highlighting potential mismatches between administrative welfare records and actual socio-economic conditions. This is particularly relevant in contexts where official social assistance databases may not always be fully updated or comprehensive. By combining clustering and ranking, the university can move beyond document-based screening toward a more nuanced evaluation that better reflects the lived realities of students.

4. CONCLUSION

This study successfully integrates the Density-Based Spatial Clustering of Applications with Noise (DBSCAN) algorithm and the Technique for Order Preference by Similarity to Ideal Solution (TOPSIS) method in the selection process of prospective recipients of the Indonesia Smart Lecture Card (KIP-K) scholarship at the University of Muhammadiyah Enrekang. The integration of these two methods allows for a more objective evaluation process by utilizing multidimensional data analysis that reflects the socioeconomic conditions and academic achievements of prospective students. The results of the study show that the DBSCAN algorithm is able to effectively group student data based on data density without requiring the determination of the number of clusters at the beginning. With optimal parameters of Eps = 0.6 and MinPts = 3, the model successfully generated 10 main clusters and one noise cluster, which illustrated the variation in the socioeconomic conditions of prospective scholarship recipients. This clustering process provides a deeper understanding of the data structure and is able to identify student groups with similar characteristics and outlier data. Furthermore, the TOPSIS method is used to rank alternatives based on the proximity to the positive ideal solution and the distance to the negative ideal solution by considering the criteria of benefit and cost. The results of the ranking showed that candidates with high levels of socioeconomic vulnerability and limited access to basic facilities tended to obtain higher eligibility scores. In addition, the integration of TOPSIS with DBSCAN clustering results allows the identification of potential candidates from both minority and outlier clusters who still have high eligibility scores. Overall, the combination of the DBSCAN-TOPSIS method is able to increase transparency, objectivity, and efficiency in the scholarship selection process compared to conventional selection methods. The resulting model also shows a selection accuracy rate of 82.4%, so it has the potential to be used as a Decision Support System (DSS) in the selection process of scholarship recipients at universities. The implementation of this model is expected to help speed up the decision-making process, reduce subjective bias, and produce more targeted recommendations for scholarship recipients. The findings of this study suggest that an approach based on the integration of clustering methods and multicriteria decision-making can be an innovative alternative in a data-driven scholarship selection system. This approach also has the potential to be applied to other educational institutions to support a fairer and more transparent selection process for educational assistance, while contributing to increasing access to higher education as part of efforts to achieve the Sustainable Development Goals (SDGs), especially in the aspect of equitable access to quality education.

ACKNOWLEDGMENT

Thank you to the Ministry of Higher Education, Science and Technology of the Republic of Indonesia for the financial support that has been provided in the national competitive grant program, Beginner Lecturer Research, so that this research can be carried out properly. Thank you also to the Muhammadiyah University of Enrekang for giving permission to carry out this research.

REFERENCES

[1] A. Ambariyanto and Y. J. Utama, "Educating higher education institutions to support SDGs: Indonesian case," in *E3S Web Conf.*, 2020, doi: 10.1051/e3sconf/202020202015.

[2] A. E. Adeyemi, J. Ahn, Z. Xu, H. Muko, and B. Matt, "Promoting SDGs through education: A theory of planned behavior analysis of Japanese and Nigerian students' sustainability actions," *Sustain. Develop.*, 2025, doi: 10.1002/sd.3493.

[3] M. Bahtilla and X. Hui, "The principal as a curriculum-instructional leader: A strategy for curriculum implementation in Cameroon secondary schools," *Int. J. Educ. Res.*, vol. 8, no. 4, 2020.

[4] F. Irhamsyah, "Sustainable Development Goals (SDGs) dan dampaknya bagi ketahanan nasional," *J. Lemhannas RI*, vol. 7, no. 2, pp. 45–54, 2020, doi: 10.55960/jlri.v7i2.71.

[5] N. Azizah, M. Marsofiyati, and E. D. Utari, “The influence of the help of the Smart Indonesia Collage Card (KIPK) on the motivation of student studying at the Faculty of Economics and Business, State University of Jakarta,” *Asian J. Appl. Educ. (AJAE)*, vol. 4, no. 3, pp. 321–334, Jul. 2025, doi: 10.55927/ajae.v4i3.14865.

[6] B. Octafiani, S. S. N. Siti, and B. B. Masitho, “Implementation of the KIP Kuliah program for aspiration path for students of Universitas Mandiri Bina Prestasi,” *J. Compr. Sci. (JCS)*, vol. 4, no. 2, pp. 697–707, 2025.

[7] I. Akbar, I. S. Samad, R. Rahmat, and S. Rosmiana, “Data mining analysis of K-means algorithm and decision tree for early detection of students at risk of dropping out,” *J. Informat. Inf. Syst. Softw. Eng. Appl. (INISTA)*, vol. 7, no. 2, pp. 148–162, 2025.

[8] S. Nagaraju, M. Kashyap, and M. Bhattachraya, “An effective density based approach to detect complex data clusters using notion of neighborhood difference,” *Int. J. Autom. Comput.*, vol. 14, no. 1, 2017, doi: 10.1007/s11633-016-1038-7.

[9] U. Rahmalisa and M. Muhardi, “Penerapan metode TOPSIS untuk seleksi penerima beasiswa (studi kasus: SMAN 2 Tebing Tinggi Timur),” *J. Teknol. Sist. Inf. Apl.*, vol. 2, no. 1, 2019, doi: 10.32493/jtsi.v2i1.2687.

[10] Z. Arifin, “A comprehensive analysis of KIP Kuliah scholarship recipients conditions in Central Java private universities,” in *Proc. Int. Conf. Sci., Educ., Technol.*, vol. 10, pp. 40–49, Sep. 2024.

[11] D. Haruna, H. A. Karim, and Adriansyah, “Analysis of education financing strategies to improve accessibility in higher education,” *ICMIE Proc.*, vol. 2, no. 1, pp. 11–19, Jul. 2025, doi: 10.30983/ICMIE.V1I1.47.

[12] R. C. A. Fajardo, F. B. Yara, R. F. Ardeña, M. K. L. Hernandez, and J. C. T. Arroyo, “A data-driven approach in predicting scholarship grants of a local government unit in the Philippines using machine learning,” *Int. J. Eng. Trends Technol.*, vol. 72, no. 6, pp. 74–81, Jun. 2024, doi: 10.14445/22315381/IJETT-V72I6P108.

[13] S. A. Asri, I. G. N. B. Caturbawa, P. M. Prihatini, N. W. Rasmini, I. M. R. A. Nugroho, and E. Rudiastari, “Scholarship application with decision support system feature using progressive web app,” pp. 330–338, Dec. 2024, doi: 10.2991/978-94-6463-587-4_38.

[14] S. Sharief, M. B. Balaji, and A. Prof, “Scholarship prediction system using machine learning,” *Int. J. Sci. Res. Eng. Develop.*, vol. 8, 2025.

[15] A. A. Bushra, D. Kim, Y. Kan, and G. Yi, “AutoSCAN: Automatic detection of DBSCAN parameters and efficient clustering of data in overlapping density regions,” *PeerJ Comput. Sci.*, vol. 10, 2024, doi: 10.7717/peerj-cs.1921.

[16] M. R. Ridho, H. Hairani, K. A. Latif, and R. Hammad, "Kombinasi metode AHP dan TOPSIS untuk rekomendasi penerima beasiswa SMK berbasis sistem pendukung keputusan," *J. Tekno Kompak*, vol. 15, no. 1, 2021, doi: 10.33365/jtk.v15i1.905.

[17] B. G. Sudarsono and S. P. Lestari, "Clustering penerima beasiswa yayasan untuk mahasiswa menggunakan metode K-means," *J. Media Informat. Budidarma*, vol. 5, no. 1, 2021, doi: 10.30865/mib.v5i1.2670.

[18] A. Iskandar, "Penerapan algoritma K-medoids untuk clustering prioritas penerima beasiswa," *J. Inf. Syst. Res. (JOSH)*, vol. 4, no. 2, 2023, doi: 10.47065/josh.v4i2.2927.

[19] W. Supriyanti, S. Kom, and M. Kom, "Machine learning: Konsep, algoritma, dan implementasi," *JURIKOM (J. Ris. Komput.)*, vol. 11, no. 3, pp. 90–108, Jun. 2025, doi: 10.30865/JURIKOM.V11I3.8462.

[20] Y. Yanto, A. Homaidi, and A. Lutfi, "Implementasi metode clustering dengan algoritma DBSCAN untuk identifikasi sentra industri berbasis Google Map," *G-Tech: J. Teknol. Terap.*, vol. 8, no. 3, pp. 2112–2121, Jul. 2024, doi: 10.33379/gtech.v8i3.4959.

[21] S. Chakraborty, P. Chatterjee, and P. P. Das, "Technique for order of preference by similarity to ideal solution (TOPSIS)," in *Multi-Criteria Decision-Making Methods in Manufacturing Environments*, pp. 85–97, Aug. 2023, doi: 10.1201/9781003377030-8.

[22] F. Sulianta, *Dasar dan Konsep Machine Learning*. 2025.

[23] Q. Zhu, X. Tang, and Z. Liu, "Revised DBSCAN clustering algorithm based on dual grid," in *Proc. Chin. Control Decis. Conf. (CCDC)*, pp. 3461–3466, Aug. 2020, doi: 10.1109/CCDC49329.2020.9163926.

[24] G. Liu, "The study of network intrusion detection based on CNN-GRU-TWD," 2024, doi: 10.1117/12.3038184.

[25] G. R. W. Syurifah, "Implementasi metode ST-DBSCAN untuk pengelompokan pola persebaran titik api pada data kebakaran hutan di Indonesia," doctoral dissertation, Univ. Islam Negeri Maulana Malik Ibrahim, 2024.

[26] I. N. Simbolon and P. D. Friskila, "Analisis dan evaluasi algoritma DBSCAN (Density-Based Spatial Clustering of Applications with Noise) pada tuberkulosis," *J. Informat. dan Tek. Elektro Terap.*, vol. 12, no. 3S1, 2024.

[27] M. A. Wijaya, D. S. Prayoga, A. K. Rahman, and A. P. Sari, "Perbandingan algoritma K-means dan DBSCAN dalam metode clustering dengan PCA untuk analisis data statistik negara dunia," in *Proc. Semin. Nas. Informat. Bela Negara (SANTIKA)*, vol. 3, pp. 63–70, Nov. 2023.

[28] R. M. Simanjorang, A. Simangunsong, A. Sitohang, J. L. Tobing, and S. Simanjorang, "Sistem pendukung keputusan pemilihan guru berprestasi dengan metode TOPSIS," *J. Media Informat.*, vol. 7, no. 1, pp. 415–426, 2026.

[29] T. Rahmadani, "Status sosial ekonomi rumah tangga dan pengetahuan gizi ibu balita di Kecamatan Tulang Bawang Tengah," doctoral dissertation, Univ. Lampung, 2024.