



Article

Application of TOPSIS and K-Means Clustering Methods in Recommendations and Analysis of Study Program Interest for New Students

Hidayati Rusnedy^{1*}, Kasini², Lailatul Syifa Tanjung³, Yesi Yusmita⁴

^{1,2}Department of Informatics Engineering, Faculty of Engineering, Pahlawan Tuanku Tambusai

^{3,4}Department of Industrial Engineering, Faculty of Engineering, Pahlawan Tuanku Tambusai

E-mail: *hidayati@perguruanunggipahlawan.ac.id, kasiniaqm@gmail.com (Corresponding author)

ARTICLE INFORMATION

Volume 5 Issue 1

Received: 17 February 2025

Accepted: 28 March 2025

Publish Online: 31 March 2025

Online: at <https://JESTM.org/>

ABSTRACT

The selection of a study program is one of the crucial initial decisions for prospective students when entering college. This decision is ideally based on a good understanding of their interests, talents, and abilities so that prospective students can study optimally and in accordance with their potential. Technique for Order Preference by Similarity to Ideal Solution (TOPSIS) is a multi-criteria decision-making method where the best alternative has the longest distance from the negative ideal solution and has the shortest distance from the positive ideal solution. The selection of these criteria and alternatives aims to produce relevant and accurate recommendations in helping prospective students determine the choice of study program that best suits their potential and preferences. The results of these recommendations are then further analyzed to group the results of the recommendations based on the category of interest. The K-Means Clustering method using the K-Means method with the results of C1 with 9 Respondentsts of less interested study programs, C2 with 25 Respondentsts of moderately interested study programs, and C3 with 21 Respondents of highly interested study programs.

Keywords

Decision Support System

Topsis

Clustering

K-Means

Pahlawan Tuanku Tambusai University

1. Introduction

Every year, 12th-grade high school students (prospective students) are faced with the decision of which study program to pursue at university. Choosing a study program is one of the crucial initial decisions prospective students make when entering college. Ideally, this decision should be based on a good understanding of their interests, talents, and abilities so that prospective students can pursue their studies optimally and in accordance with their potential. However, in reality, many prospective students experience difficulties in making this choice due to limited information and a lack of understanding of the study programs available at universities (Santoso, 2025).

Confusion and anxiety in choosing a study program can influence prospective students' decision-making. Study program selection is often based on the influence of friends and parents' wishes, without considering important aspects such as individual interests, talents, and abilities (Renatalia et al., 2020). This often leads to further impacts such as low motivation to study, declining academic performance, and the desire to change study program. Therefore, a decision support system is needed to assist prospective students in determining the study program that best aligns with their interests, talents, and abilities.

Decision Support System (DSS) is a computer-based system designed to assist in complex, semi-structured, or unstructured decision-making processes. This system uses a multi-criteria approach to generate alternative recommendations as solutions to problems encountered (Okfalisa et al., 2020).

The Technique for Order Preference by Similarity to Ideal Solution (TOPSIS) is a multi-criteria decision-making method in which the best alternative has the longest distance from the negative ideal solution and the shortest distance from the positive ideal solution (Krohling & Pacheco, 2015; Hts et al., 2024). The TOPSIS method was chosen because of its several advantages, including its ease of understanding and application. Furthermore, this method can mathematically assess the relative performance of various decision alternatives using a simple yet effective approach (Sari et al., 2018).

In processing the data using the TOPSIS method, this study used three main criteria: interest, talent, and ability. These three criteria were chosen because they are considered important factors in determining the suitability of prospective students for study programs. A total of 19 alternatives were analyzed, representing the 19 undergraduate study programs available at Pahlawan Tuanku Tambusai University. The selection of these criteria and alternatives aims to generate relevant and accurate recommendations to help prospective students determine the study program that best aligns with their potential and preferences.

In addition to generating recommendations, it is also crucial for higher education institutions to understand the overall trends in study program interest. Therefore, further analysis is needed to group the recommendations based on interest categories. The K-Means Clustering method is an effective clustering technique for identifying patterns and segmenting data. This algorithm seeks well-defined partitions so that the squared distances of the clusters within them are (Wang et al., 2019). Through this grouping, study programs can be classified into high, medium, and low interest categories, allowing universities to conduct strategic evaluations and planning in managing the study programs offered.

In this study, the recommendations will be grouped into three clusters: Cluster 1 (C1) for low-interest study programs, Cluster 2 (C2) for medium-interest study programs, and Cluster 3 (C3) for high-interest study programs. By combining the TOPSIS method in providing study program recommendations and K-Means Clustering in cluster analysis, this study aims to provide a more comprehensive solution in selecting study programs and analyzing new student interests. This research is expected to help students in determining the study program that best suits their interests, talents, and abilities, as well as assist universities in understanding study program selection trends based on existing data.

2. Literature Riview

2.1. Technique for Order by Similarity to Ideal Solution (TOPSIS)

Among the multicriteria methods, the method that is widely used and proven to be effective in various applications is the Order Preference Technique by Similarity to Ideal Solution. (TOPSIS) (Radulescu et al., 2024). TOPSIS is a popular multicriteria method known for its simplicity, strong mathematical foundation, and ease of application in multicriteria problems (Chakraborty, 2022). The following are the stages of the TOPSIS method (Okfalisa et al., 2022):

1. Determining the normalized decision matrix (r_{ij}).

$$r_{ij} = \frac{x_{ij}}{\sqrt{\sum_{i=1}^m x_{ij}^2}} \quad (1)$$

2. Calculating the weighted normalized decision matrix. The positive ideal solution $+$ and the negative ideal solution can be determined based on the normalized weight rating (y_{ij}) as.

$$y_{ij} = w_i r_{ij} \quad (2)$$

3. Calculating the positive ideal solution matrix and the negative ideal solution matrix.

$$A^+ = (y_1^+, y_2^+, \dots, y_n^+) \quad (3)$$

$$A^- = (y_1^-, y_2^-, \dots, y_n^-) \quad (4)$$

Assumption:

$$y_j^+ = \begin{cases} \max y_{ij} & \text{if } j \text{ is a benefit attribute} \\ \min y_{ij} & \text{if } j \text{ is the attribute cost} \end{cases}$$

$$y_j^- = \begin{cases} \max y_{ij} & \text{if } j \text{ is a benefit attribute} \\ \min y_{ij} & \text{if } j \text{ is the attribute cost} \end{cases} \quad (5)$$

4. Calculating the distance (d) between the values of each alternative with the positive ideal solution matrix (d^*) and the negative ideal solution matrix (d).

$$d_1^* = \sqrt{\sum_{j=1}^n (v_{ij} - v_j^*)^2}, i=1, \dots, m \quad (6)$$

$$d_1^- = \sqrt{\sum_{j=1}^n (v_{ij} - v_j^-)^2}, i=1, \dots, m \quad (7)$$

5. Calculating preference values for each alternative (CC).

$$CC_i = \frac{d_i^-}{d_i^* + d_i^-}, i=1, \dots, m \quad (8)$$

2.2. K-Means

K-Means is one of the simplest and most widely used clustering algorithms, often used as a standard method in various machine learning studies. The primary goal of this method is to identify centroids and minimize the total squared distance between each data point and the nearest cluster center. (Nie et al., 2023).

The main function of the K-Means algorithm is to divide data into a number of K mutually exclusive clusters, where each cluster is formed around a single cluster center (centroid) calculated iteratively. This process aims to reduce variation within each cluster by positioning the centroid as close as possible to the surrounding data points compared to the centers of other clusters. In this clustering, a measure of similarity between data—generally calculated using the Euclidean distance—plays an important role in determining the proximity of data to a centroid, although in some cases other distance measures can be used (Miraftabzadeh et al., 2023). The following are the steps of the K-Means algorithm (Kasini et al., 2025):

1. Determine the number of k values as the number of clusters.
2. Allocate data into groups randomly.
3. Calculate the cluster center (centroid) using the mean for each cluster.

$$D_{(i,j)} = \sqrt{(X_{1i} - X_{1j})^2 + (X_{2i} - X_{2j})^2 + \dots + (X_{ki} - X_{kj})^2} \quad (9)$$

Where:

$D_{(i,j)}$ = Distance of data point i to cluster center j

X_{ki} = Data point i on data attribute k

X_{kj} = Center point j on attribute k

4. Allocate data based on the closest distance between the data and its centroid.

5. Return to the previous step, if it turns out that there is still data that has moved clusters or if the centroid value is above the threshold value, or if the value in the objective function used is still above the threshold.

3. Research Methodology

This section systematically explains the stages carried out in the research that focuses on the application of the TOPSIS and K-Means Clustering methods in the recommendation and analysis of study program interests of new students. The research stages begin with the collection of primary data through the distribution of questionnaires, then the data obtained is processed using the Technique for Order Preference by Similarity to Ideal Solution (TOPSIS) method to produce recommendations for study programs that best suit the characteristics of each student. After the recommendations are obtained, further analysis is carried out using the K-Means Clustering method to group study programs based on the number of interested people into three categories: high, medium, and low interest. This stage aims to obtain an overview of the trends in interest of prospective new students.

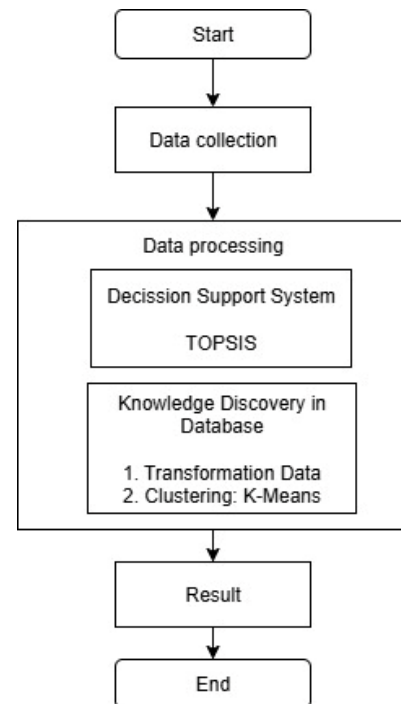


Figure 1 Research Methodology

Before data processing using the TOPSIS method to generate study program recommendations and the K-Means method for clustering, the initial stage was data collection by distributing questionnaires to 55 12th-grade high school students or equivalent. The questionnaire was designed to measure students' interests, talents, and abilities in various fields of study. The questionnaire was chosen because it was considered effective for directly reaching respondents and obtaining primary data relevant to the research objectives.

After obtaining the study program recommendations using the TOPSIS method, further analysis was conducted to identify overall interest patterns using the K-Means Clustering method. The purpose of this analysis was to group study programs based on the number of students receiving recommendations into three categories: Cluster 1 (high interest), Cluster 2 (moderate interest), and Cluster 3 (low interest).

interest). The results of this process provide an overview of the distribution of prospective student interest in available study programs and can serve as a strategic reference for universities in terms of promotion, curriculum development, and more effective study program capacity planning.

4. Result and Discussion

This section presents the key findings of the study in a structured and logical manner, followed by interpretation and discussion in relation to the research objectives, relevant theories, and previous studies. Authors are encouraged to use clear tables and figures to support the presentation of results.

4.1. Recommendation Result

Relevant study program recommendations were obtained using the TOPSIS method, using 3 assessment criteria (Table 1) and 19 alternatives from all undergraduate study programs at Pahlawan Tuanku Tambusai University (Table 2). Each criterion is weighted to reflect its level of importance in decision-making. Table 1 shows that each criterion is categorized into benefit categories, with a preference weighting of 0.40 (40%), 0.35 (35%) for Ability, and 0.25 (25%) for Interest. In the TOPSIS method, these weights and criteria are used as the main parameters to determine the best alternative based on the highest preference value (V).

Table 1. Criteria

Criteria	Category	Preference Weight
Interest	Benefit	0.25
Talent	Benefit	0.35
Ability	Benefit	0.40

19 undergraduate study programs available at Pahlawan Tuanku Tambusai University were used as alternatives in the TOPSIS decision-making process. These programs include Gizi, Kesehatan Masyarakat, Keperawatan, Kebidanan, Pendidikan Guru Sekolah Dasar (PGSD), Pendidikan Guru Pendidikan Anak Usia Dini (PG PAUD), Pendidikan Matematika, Pendidikan Bahasa Inggris, Pendidikan Jasmani, Teknik Informatika, Teknik Sipil, Teknik Industri, Hukum, Peternakan, Biologi, Kewirausahaan, Bisnis Digital, Ekonomi Syariah, and Perbankan Syariah. These alternatives served as the basis for determining the recommended study programs that best meet the established criteria.

Based on the calculation results using the TOPSIS method, a final weighting was obtained that reflects the level of suitability of each study program (prodi) with the preferences or interests, talents, and abilities of each Respondentst. The weighting value ranges from 0 to 1, where a value closer to 1 indicates that the study program is more in line with the selection criteria set by the Respondentst. The results of the study program recommendations, including the final weighting value and selected alternatives, can be seen in Table 2.

Table 2. Study Program Recommendation Results for Respondentst 1

Study Program	Weight	Ranking
Gizi	0.486	16
Kesehatan Masyarakat	0.750	2
Keperawatan	0.711	6
Kebidanan	0.826	1
PGSD	0.467	17
PG PAUD	0.732	4
Pendidikan Matematika	0.547	14
Pendidikan Bahasa Inggris	0.605	12
Penjaskesrek	0.669	8
Teknik Informatika	0.590	13
Teknik Sipil	0.510	15
Teknik Industri	0.711	5
Hukum	0.618	11
Kewirausahaan	0.734	3
Bisnis Digital	0.464	18
Peternakan	0.634	9
Biologi	0.619	10
Ekonomi Syariah	0.419	19
Perbankan Syariah	0.684	7

The results of recommendations from all Respondentsts can be seen in Table 4, which presents a summary of the results of recommendations for selecting each study program based on the results of the TOPSIS method analysis.

Table 3. Recapitulation of Study Program Selection Recommendation Results

Respondentsts	SelectedStudy Programs	Final Weight
Respondents 1	Kebidanan	0.826
Respondents 2	PG PAUD	0.924
Respondents 3	PGSD	1.000
Respondents 4	Kewirausahaan	0.826
Respondents 5	Perbankan Syariah	0.830
...	...	...
...	...	...
...	...	...
Respondents 50	Kesehatan Masyarakat	0.835
Respondents 51	Teknik Industri	0.805
Respondents 52	Pendidikan Matematika	1.000
Respondents 53	Pendidikan Bahasa Inggris	0.694
Respondents 54	Pendidikan Bahasa Inggris	0.695
Respondents 55	Teknik Industri	0.809

4.2. Clustering Results

The initial data used in the clustering process using the K-Means method comes from the results of study program recommendations using the TOPSIS method. The data is then transformed into a numeric format as shown in Figure 1. This transformation is carried out to adapt the data format to the needs of the K-Means algorithm, which requires input data in numeric form.

Responden	Prodi Terpilih	Fakultas
Responden 1	Keluhan	Fakultas Ilmu Kesehatan
Responden 2	PG PAUD	Fakultas Keguruan dan Ilmu Pendidikan
Responden 3	PGSD	Fakultas Keguruan dan Ilmu Pendidikan
Responden 4	Kewirausahaan	Fakultas Ekonomi dan Bisnis
Responden 5	Perbankan Syariah	Fakultas Agama Islam
...	...	...
Responden 50	Kesehatan Masyarakat	Fakultas Ilmu Kesehatan
Responden 51	Teknik Industri	Fakultas Teknik
Responden 52	Pendidikan Matematika	Fakultas Keguruan dan Ilmu Pendidikan
Responden 53	Pendidikan Bahasa Inggris	Fakultas Keguruan dan Ilmu Pendidikan
Responden 54	Pendidikan Bahasa Inggris	Fakultas Keguruan dan Ilmu Pendidikan
Responden 55	Teknik Industri	Fakultas Teknik

Figure 2. Transformasi Data

Furthermore, clustering data processing using the K-means method was carried out 5 iterations. Based on the iteration results, study programs were grouped into three clusters based on their level of interest. Cluster 1, which consists of Kewirausahaan, Perbankan Syariah, Peternakan, Bisnis Digital, dan Biologi study programs, is categorized as a study program with low interest with a total of 9 Respondents. Cluster 2 includes Teknik Informatika, Teknik Industri, Pendidikan Jasmani, Pendidikan Bahasa Inggris, Pendidikan Matematika, dan Hukum study programs, which are classified as study programs with a moderate level of interest with a total of 25 Respondents. Meanwhile, Cluster 3 contains Midwifery, PG PAUD, PGSD, Nutrition, Nursing, and Public Health study programs, which are classified as study programs with high interest, with a total of 21 Respondents. The Iteration Results can be seen in Figure 3.

No	Responden	Prodi Terpilih	Fakultas	C1	C2	C3	Jarak Terpendek	Cluster
1	Responden 1	4	1	13	6	1	1	3
2	Responden 2	6	2	11	4	2	2	3
3	Responden 3	5	2	12	5	1	1	3
4	Responden 4	14	6	2	5	11	2	1
5	Responden 5	19	7	3	10	16	3	1
...	...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...	...
50	Responden 50	2	1	15	8	2	2	3
51	Responden 51	12	3	5	2	8	2	2
52	Responden 52	7	2	10	3	3	3	3
53	Responden 53	8	2	9	2	4	2	2
54	Responden 54	8	2	9	2	4	2	2
55	Responden 55	12	3	5	2	8	2	2

Figure 3. Clustering Result

5. Conclusion

Based on the results of research using the Topsis and K-Means methods at Pahlawan Tuanku Tambusai University, it can be concluded that the Topsis method provides recommendations for study programs for high school student in choosing a study programs according to their interest indicators. From the results of these recommendations, they are then processed using the K-Means method with the

results of C1 with 9 Respondentsts of less interested study programs, C2 with moderate interest totaling 25 Respondentsts, and C3 with 21 Respondentsts of highly interested study programs. At Pahlawan Tuanku Tambusai University, there are still limited resources in implementing study programs promotions at school, this can be seen by the fact that many prospective students still do not know the profile and advantages of each existing study programs. With recommendations using the Topsis method and Clustering using K-Means, it is hoped that there will be increased socialization to see the interests of prospective students and understand the study programs to be chosen.

References

- Chakraborty, S. (2022). TOPSIS and Modified TOPSIS: A comparative analysis. *Decision Analytics Journal*, 2(September 2021), 100021.
- Hts, D. I. G., Desi, E., Aliyah, S., Nasution, F. P., Indriani, U., & Edi, F. (2024). Decision Support System Using the TOPSIS Method in New Teacher Selection. *Sinkron*, 8(3), 1706–1714. <https://doi.org/10.33395/sinkron.v8i3.13751>
- Kasini, Rusnedy, H., Tanjung, L. S., & Munti, N. Y. S. (2025). Penerapan algoritma k-means untuk pengelompokan data mahasiswa baru Program Studi Teknik Informatika di Universitas Pahlawan Tuanku Tambusai. *JUTIN : Jurnal Teknik Industri Terintegrasi*, 8(1).
- Krohling, R. A., & Pacheco, A. G. C. (2015). A-TOPSIS - An approach based on TOPSIS for ranking evolutionary algorithms. *Procedia Computer Science*, 55(Itqm), 308–317. <https://doi.org/10.1016/j.procs.2015.07.054>
- Miraftabzadeh, S. M., Colombo, C. G., Longo, M., & Foidadelli, F. (2023). K-Means and Alternative Clustering Methods in Modern Power Systems. *IEEE Access*, 11(September), 119596–119633. <https://doi.org/10.1109/ACCESS.2023.3327640>
- Nie, F., Li, Z., Wang, R., & Li, X. (2023). An Effective and Efficient Algorithm for K-Means Clustering With New Formulation. *IEEE Transactions on Knowledge and Data Engineering*, 35(4), 3433–3443. <https://doi.org/10.1109/TKDE.2022.3155450>
- Okfalisa, O., Rusnedy, H., Iswavigra, D. U., Pranggono, B., Haerani, E. H., & Saktioto, S. (2020). Decision Support System for Smartphone Recommendation: the Comparison of Fuzzy Ahp and Fuzzy Anp in Multi-Attribute Decision Making. *Sinergi*, 25(1), 101. <https://doi.org/10.22441/sinergi.2021.1.013>

- Okfalisa, Siburian, R., Vitriani, Y., Rusnedy, H., Saktioto, & Yola, M. (2022). Job Training Recommendation System: Integrated Fuzzy AHP and TOPSIS Approach BT - Advances on Intelligent Informatics and Computing (F. Saeed, F. Mohammed, & F. Ghaleb (eds.); pp. 84–94). Springer International Publishing.
- Radulescu, C. Z., Radulescu, M., & Boncea, R. (2024). A Linear Trade-off Group TOPSIS method with application for Internet of Things devices ranking. *Procedia Computer Science*, 242, 528–535.
<https://doi.org/10.1016/j.procs.2024.08.099>
- Renatalia, Y., Asfi, M., & Fahrudin, R. (2020). Sistem Pendukung Keputusan Pemilihan Program Studi Menggunakan Metode Profil Matching. *JURNAL DIGIT*, 10(2), 148–160.
- Santoso, N. Y. (2025). Analisis Perbandingan K- Nearest Neighbors dan Naive Bayes Untuk Rekomendasi Pilihan Program Studi Bagi Mahasiswa. *Idealis: Indonesia Journal Information System*, 8(1), 117–126.
- Sari, D. R., Windarto, A. P., Hartama, D., & Solikhun, S. (2018). Decision Support System for Thesis Graduation Recommendation Using AHP-TOPSIS Method. *Jurnal Teknologi Dan Sistem Komputer*, 6(1), 1–6.
<https://doi.org/10.14710/jtsiskom.6.1.2018.1-6>
- Wang, S., Li, M., Hu, N., Zhu, E., Hu, J., Liu, X., & Yin, J. (2019). K-Means Clustering With Incomplete Data. *IEEE Access*, 7, 69162–69171.
<https://doi.org/10.1109/ACCESS.2019.2910287>