khazanah informatika

Jurnal Ilmu Komputer dan Informatika

Combination of K-Means and Simple Additive Weighting in Deciding Locations and Strategies of University Marketing

Muhamad Ali Kasri*, Handaru Jati
Postgraduate Study in Electronics and Informatics Education
Universitas Negeri Yogyakarta
Yogyakarta, Indonesia
*Correspondence: muhamadalikasri@gmail.com

Abstract-Every year UNIMUDA Sorong welcomes new students and keeps promoting to attract more. The process generates a growing number of student data. On the other hand, the promotional strategy to attract new students faces obstacles such as generalization among locations, ineffective time, limited personnel to carry out promotions, and cost inefficiency. This study examines the new student data and university marketing strategies to optimize time, effort, and cost. It uses the K-Means method for data grouping and the Simple Additive Weighting (SAW) for ranking the results of data grouping. The result of this research suggests that location of promotion may be determined from the clustering process using the K-Means method. The silhouette coefficient test invalidates the data clustering, and the SAW method helps the ranking process to obtain a sequence of promotion locations. The ranking results reflect the predetermined decision table that directs promotion location selection according to the promotion strategy. The combination of the two methods helps to decide the location and marketing strategy to optimize time, effort, and cost. The results of this study may be used as a comparative reference for the management to decide the right promotion strategy based on the locations and student background.

Keywords: k-means, simple additive weighting, promotion, new students

Article info: submitted June 25, 2020, revised August 10, 2020, accepted Sept 30, 2020

1. Introduction

Competition to attract new student is very tight. Both public and private universities spend much money on promotion by advertising advantages such as low tuition fees and the ease of getting a job after graduation [1]. To survive, education institutions need to arrange the right promotion. Determining the right strategies may reduce costs and achieve promotional goals [2]. The community customers of educational services are very critical in questioning and selecting quality educational institutions. Parents are worried that their children may not be able to compete in the job market. [3].

Universitas Pendidikan Muhammadiyah (UNIMUDA) Sorong has an annual agenda of promotion to attracts new students within and outside the province of West Papua. Determining the location of the promotional activity refers to habitual practices and intuition every year, so it is necessary to investigate whether the location determination is right on target. Determining location quickly and accurately is not an easy task. Many things need to be considered, such as time, effort, and cost [1]. So far, the promotion strategy does not consider the target location. Using the same promotion strategy for all places may result in time-inefficiency, arduous effort, and high costs. Promotional activities have not used student data as a reference in selecting the methods and promotion strategies in attracting new students.

The student data can be analyzed using data mining techniques [4]. Such techniques process student admission data from past years to find patterns or information [5]. Patterns or information may be useful as a reference for making marketing policies [6]. Policies made based on student data are very important in streamlining limited time, energy, and costs.

The use of student data to make decisions regarding new student promotions has been widely observed. Rusli et al. have examined the decision support system for determining the location of promotion using the Analytical Hierarchy Process (AHP) and Technique Order Preference by Similarity to Ideal Solution (TOPSIS). This effort was made to improve the method of determining random promotional locations so that it becomes targeted, namely determining school locations that have the potential to become promotional sites. The data criteria observed were districts and schools. The AHP and TOPSIS methods are combined in the first and second stage selection processes. The decision support system that was built was evaluated with a User Acceptance Test for further implementation [7].

Other observations were made using the AHP fuzzy method to determine potential areas as higher education promotion targets. Determining the location using last year's registrant data is the problem raised and solved using the AHP fuzzy method with the Triangular Fuzzy Number approach. The data criteria were distance, number of schools, number of students and last year's registrants [8]. Research on campus promotion strategy policies has been carried out using the Weighted Aggregated Sum Product Assessment (WASPAS) method. The problem raised was the random selection of promotion locations so that it was unknown which school had the most potential [9].

Zanakis et al. seeks to compare many decision-making methods such as Elimination Et Choix Traduisant la Realite (ELECTRE), TOPSIS, Multiplicative Exponential Weighting, and Simple Additive Weighting (SAW) Methods. Their observations showed that the best performance was produced by the SAW and MEW methods, followed by TOPSIS, AHP and ELECTRE [10]. Meanwhile, research conducted by Sunarti, namely Comparison of TOPSIS and SAW Methods For Home Selection, concluded that the calculation of the SAW method is more recommended than the TOPSIS method [11]. The comparison between the AHP and SAW methods conducted by Wicaksono shows that the SAW method has the highest accuracy compared to the AHP method [12]. Meanwhile, the comparison between the AHP and SAW fuzzy methods conducted by Erdiansyah et al obtained the results that SAW is more recommended than fuzzy AHP [13].

Studies that are described in the above paragraphs and related to the problems faced have led this study to use the SAW method in determining the location and strategy of university marketing to attract new students. However, Korsemov and Borrisova stated that the SAW method would be better and work more optimally with the help of other methods [14]. Therefore, this study combines SAW with the K-means. The reason is that the K-means method uses non-abstract and clear physical data, which is suitable for the data used in this study. Besides, the K-means method can classify a large number of data with relatively fast and efficient computation time [15].

Jain et al. Described the use of K-means as a clustering method which has been quite reliable for a long time [16]. The development of K-means has undergone several modifications with the emergence of variations such as Fuzzy C-means. However, Velmurugan revealed that the K-means algorithm performs better than the Fuzzy C-means algorithm [17]. In addition, in a survey published by Springer "Top 10 Algorithms in Data Mining", the K-means algorithm is placed in position 2 (two) as the most widely used algorithm in data mining and in the first position for clustering algorithms [18].

The contribution of this paper is to describe the research that combines the K-means method with the SAW method in determining the target location for promotion related to promotional strategies applied. The K-means method is used to divide student data into several clusters where each cluster has a different priority level in determining the location and strategy. The data for each cluster is further processed by the SAW method to obtain the final result in the form of a ranking. The results of this ranking become recommendations for decision makers in determining marketing strategies to attract new students.

2. Method

Combining the two methods can be done if one method functions as a data divider into several clusters and the other method acts as an alternative ranking. In this case, the K-means method is used to determine the best cluster. The best clusters are locations that match the predetermined data criteria. Meanwhile, the SAW method is used for the alternative data ranking process. The results of this ranking are used as a reference for recommendations in promoting new students at the selected location.

a. Data collection

The research data were obtained directly from the source (primary data), namely from the UNIMUDA Sorong admissions office. Determination of the weight value of each data criterion used is based on the decision of the college itself.

b. Research Stage

The research stage began with determining the problem and selecting the data criteria. Furthermore, the analysis was carried out using the k-means clustering method and ranking using the SAW method. The final stage is a recommendation and conclusion (figure 1).

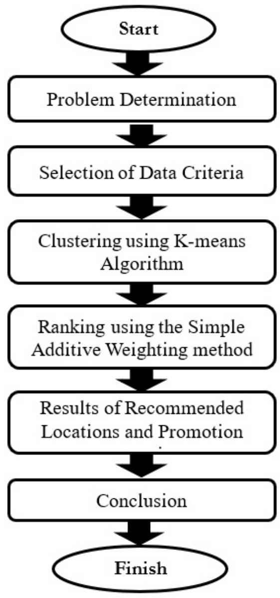

Figure 1. Research Stage

c. Problem Determination and Selection Criteria

The targets to be achieved in this research are to get the location and promotional strategy. The data criteria used in student data are sub-district data, parent's occupation, and the chosen study program.

There are arguments behind the selection of data criteria. The kecamatan data represent the distance between student residences and tertiary institutions. Residence distance is an important factor to streamline existing resources in visiting promotional locations. Parents' job data represent the economic conditions of students to examine the suitability of the strategy of promotion. Study program data was chosen because it is the place where students study. Higher education institutions focus on promotion in eastern Indonesia in accordance with government directives to advance education in the region [19].

To facilitate the data mining process, data criteria were initialized. Data initialization for each criterion can be seen in Table 1.

Table 1. Data criteria

Criteria	Sub Criteria	Value
Sub-District (V1)	1-100 km	8
	101-200 km	7
	201-300 km	6
	301-400 km	5
	401-500 km	4
	>500	3
Parents' job **(V2)**	Farmers / Fishermen	8
	General employees	7
	Government officer (either civil or military)	6
	Labor	5
	Entrepreneur	4
	Others	3
Study program **(V3)**	Agribusiness	2
	Aquaculture	2
	Indonesian Language Education	1
	English language education	1
	Biology Education	1
	Pharmacy	2
	Law	2
	Science education	2
	Mathematics education	1
	Primary teacher Education	1
	Physical education, health and recreation	1
	Civic education	1
	Information Technology Education	2
	Chemical Engineering	2
	Civil Engineering	2

The determination of the value of each sub-criterion is a decision of the college admissions team. The value of the sub-district data initialization uses the unit distance of kilometers (km) with a data range every 100 km. Selection of ranges is linked to the geography and demographics of West Papua. Travel to distances over 100 km takes a few days so it is not appropriate to lump them together in groups with distances under 100 km.

Initialization of parent job data uses a value of 1 to 8 according to the number of parents of students who are involved in the job. A score of 8 is given for the jobs most parents are involved in. Lower marks are assigned to jobs that are less engaged.

Study program data were initialized with a value of 1 and 2. A score of 1 was given to study programs that had been around for a long time, namely those that had been established in the range 2004 to 2014. While study programs that were established after 2015 were given a score of 2.

The promotion strategies used are presented in table 2, as has been carried out so far by universities.

Table 2. Promotion strategies

Code	Promotion
P1	Social media
P2	TV, Radio, Newspaper Ads - Mass media
P3	Brochure Spread
P4	Installation of banners, flyers, banners
P5	Education Exhibition
P6	Outreach to schools
P7	Scholarship
P8	Invitation Line Promotion
P9	Broadcast SMS

Recommendations on the decision tree can be seen in table 3. The decision tree is extracted from the results of interviews from the PMB committee, which are then compiled in the form of new student promotion provisions. The decision tree has determined the value to be achieved, for example if the criteria V1, V2, and V3 get values 7-8, 8, and 1, then the promotion strategies recommended is P3, P4, P5, P6.

The distance factor greatly influences the promotion strategy carried out, because of the geography of the West Papua region which consists of islands [20]. So that transportation from universities to the target location for promotion must use the sea route. Parents' work also influences the promotion strategy. Most of the jobs of parents fall into the lower middle income category, for example jobs in the agriculture, plantation, forestry, hunting, fishery and labor sectors [21]. So the promotion strategies carried out have been adjusted to the economic conditions of the students.

Table 3. Decision Tree

sub-District		Parents' job	Study program	Promotion Code
Min	*Max*			
7	8	8	1	P3,P4,P5,P6
5	6	8	1	P1,P2,P3,P5,P8
3	4	8	1	P1,P2,P9,P8
7	8	8	2	P3,P4,P5,P6,P7
5	6	8	2	P1,P2,P3,P5,P8,P7
3	4	8	2	P1,P2,P9,P8,P7
7	8	7	1	P3,P4,P5,P6
5	6	7	1	P1,P2,P3,P5,P8
3	4	7	1	P1,P2,P9,P8
7	8	7	2	P3,P4,P5,P6,P7
5	6	7	2	P1,P2,P3,P5,P8,P7
3	4	7	2	P1,P2,P9,P8,P7
7	8	6	1	P3,P4,P5,P6
5	6	6	1	P1,P2,P3,P5,P8
3	4	6	1	P1,P2,P9,P8
7	8	6	2	P3,P4,P5,P6,P7
5	6	6	2	P1,P2,P3,P5,P8,P7
3	4	6	2	P1,P2,P9,P8,P7
7	8	5	1	P3,P4,P5,P6
5	6	5	1	P1,P2,P3,P5,P8
3	4	5	1	P1,P2,P9,P8
7	8	5	2	P3,P4,P5,P6,P7
5	6	5	2	P1,P2,P3,P5,P8,P7
3	4	5	2	P1,P2,P9,P8,P7
7	8	4	1	P3,P4,P5,P6
5	6	4	1	P1,P2,P3,P5,P8
3	4	4	1	P1,P2,P9,P8
7	8	4	2	P3,P4,P5,P6,P7
5	6	4	2	P1,P2,P3,P5,P8,P7
3	4	4	2	P1,P2,P9,P8,P7
7	8	3	1	P3,P4,P5,P6
5	6	3	1	P1,P2,P3,P5,P8
3	4	3	1	P1,P2,P9,P8
7	8	3	2	P3,P4,P5,P6,P7
5	6	3	2	P1,P2,P3,P5,P8,P7
3	4	3	2	P1,P2,P9,P8,P7

d. K-means Clustering

K-means clustering is a non-hierarchical data grouping method that classifies data in the form of one or more clusters. Data that have the same characteristics are grouped into one group and data that have different characteristics are separated into other groups. Data within one group has a small degree of variation [22]. The stages of the clustering process using the K-means method are as follows [23].

1) Determine the number of clusters.
2) Select the initial centroid randomly according to the number of clusters.
3) Calculate the data distance to the centroid with the euclidean distance formula.

$$d_{xy} = \sqrt{\sum_i^n = 1 (x_i - y_i)^2} \quad (1)$$

4) Update the centroid by calculating the average value of the values in each cluster
5) Return to stage 3 if there is still data that has moved clusters or changes in the centroid value.

The determination of the K value uses the silhouette coefficient (SC) method. SC is used to see the quality and strength of the cluster, whether the objects placed in the cluster are called good. This method is used for cluster validation by combining the values of cohesion and separation. The SC value is in the range 1 to -1 with a value close to 1 which means minimizing the distance between objects in a cluster as well as maximizing the distance between the clusters. The SC value of an object i is calculated using equation (2) {24].

$$s(i)\frac{b(i)-a(i)}{\max\{(a),(b)\}} \quad (2)$$

where :
a(i) is the average distance between object i and all objects in the same cluster.
b(i) is the average distance between object i and all objects in the closest cluster.
The criteria for SC value according to Kaufman and Rousseeuw [25] are:

a) 0.7 < SI <= 1 strong structure
b) 0.5 < SI <= 0.7 medium structure
c) 0.25 < SI <= 0.5 weak structure
d) SI <= 0.25 no structure

e. Simple Additive Weighting

Simple Additive Weighting is one of the Multi Attribute Decision Making (MADM) methods [26]. This method is often known as weighted summation method [27]. The total score for an alternative is obtained by summarizing the multiplication result between the rank and the weight of each attribute [28]. The SAW method requires a decision matrix normalization process (X) to be proportional to all existing alternative ratings.
The stages of the SAW method are as follows:

1) The normalization of the matrix is adjusted according to the type of attribute so that a normalized matrix is obtained [29]. The matrix normalization calculation is shown in Equation (3).

$$r_{ij} = \begin{cases} \dfrac{x_{ij}}{max_i x_{ij}} & (benefit) \\ \dfrac{min_i x_{ij}}{x_{ij}} & (cost) \end{cases} \quad (3)$$

where :
r_{ij} : The performance rating value is normalized

x_{ij} : The attribute value that each criterion has
$x_{ij}/max_i x_{ij}$: The greatest value of each criterion
x_{ij}/min_{ij} : The smallest value for each criterion

2) The calculation of the last alternative value is done using equation (4) [30].

$$V_i = \sum_j^n = 1\, w_j r_{ij} \tag{4}$$

Information :
v_i : The final value of the alternatives
w_j : The weight value of each criterion
r_{ij} : The performance ranking value is normalized
n : Number of criteria

3) Perform ranking based on the preference value of each alternative that is the best result of the assessment.

Table 4. Weight of the Assessment Criteria

	V1	V2	V3
Attribute	Benefit	Benefit	Cost

The value of benefit and cost in equation (4) can be seen in Table 4. Determining v1 and v2 as benefit attributes because they are the main factors that affect distance and student economic conditions which in turn affect the location and promotion strategy. Meanwhile, v3 is a cost attribute as a supporting factor in sorting the distribution of students for each study program.

3. Result

This study uses student data from 2015 to 2019 with a total of 2576 people.

a. Preprocessing Data

Data preprocessing is carried out so that there is no data duplication, no missing values so that the data can be processed and does not damage the research results. The data preprocessing stages are:

1. Perform data selection.
   This is the selection of the data criteria used in the study. The results of data selection are sub-district, parent's occupation, and study program.
2. Perform data cleaning.
   The purpose of cleaning data is to remove noise, clean data that does not match the specified value, and clean empty data. So that the results of data cleaning obtained data of 2,396 students. The reduction in this data is due to the fact that some sub-district data and parents' occupations are empty, the sub-district data does not match the city / regency and province, the double occupation data for parents.
3. Transformation
   Changes in the form or initialization of research data in order to facilitate the data mining process. The provisions for the initialization value can be seen in table 1, while the results of the initialization of student data can be seen in table 5.

Table 5. Student Data

Student	V1	V2	V3
Student 1	8	8	2
Student 2	8	4	2
Student 3	8	5	2
Student 4	8	7	2
Student 5	8	6	2
Student 6	8	8	2
Student 7	8	4	2
Student 8	8	6	2
Student 9	8	8	2
...	...	...	...
Student 2396	8	7	1

Source: primary data

b. K-means calculations

The calculation is done is to determine the centroid initial randomly selected from all attribute data. Selection of the initial centroid, namely the value of k = 6. This provision is based on the initial target of promotion carried out by the admission team, namely 6 locations. This is done by considering the limited resources available to carry out promotions at the 6 locations. The initial centroid can be seen in table 6.

Table 6. Initial Centroid

Centroid	Student	V1	V2	V3
C1	Student 43	3	8	2
C2	Student 254	8	3	2
C3	Student 872	6	6	2
C4	Student 1229	5	8	2
C5	Student 1467	7	8	1
C6	Student 2193	4	6	1

Calculation of the distance of each data to each centroid is done by equation (1).

$$C1 = \sqrt{(v1 - C1v1)^2 + (v2 - C1v2)^2 + (v3 - C1v3)^2}$$
$$C1 = \sqrt{(8-3)^2 + (8-8)^2 + (2-2)^2} = 5$$
$$C2 = \sqrt{(v1 - C2v1)^2 + (v2 - C2v2)^2 + (v3 - C2v3)^2}$$
$$C2 = \sqrt{(8-8)^2 + (8-3)^2 + (2-2)^2} = 5$$
$$C3 = \sqrt{(v1 - C3v1)^2 + (v2 - C3v2)^2 + (v3 - C3v3)^2}$$
$$C3 = \sqrt{(8-6)^2 + (8-6)^2 + (2-6)^2} = 2{,}8284$$
$$C4 = \sqrt{(v1 - C4v1)^2 + (v2 - C4v2)^2 + (v3 - C4v3)^2}$$
$$C4 = \sqrt{(8-5)^2 + (8-8)^2 + (2-1)^2} = 3$$
$$C5 = \sqrt{(v1 - C5v1)^2 + (v2 - C5v2)^2 + (v3 - C5v3)^2}$$
$$C5 = \sqrt{(8-7)^2 + (8-8)^2 + (2-1)^2} = 1{,}414$$
$$C6 = \sqrt{(v1 - C6v1)^2 + (v2 - C6v2)^2 + (v3 - C6v3)^2}$$
$$C6 = \sqrt{(8-4)^2 + (8-6)^2 + (2-1)^2} = 4{,}5826$$

The calculation continues until the nth data is 2396th data. In order to obtain the distance matrix, namely c1, c2, c3, c4, c5 and c6. Comparison and selection of the closest distance between the data and the cluster center were conducted. This distance indicates that the data is in one group with the closest cluster. The results of calculating the distance of each data to each centroid in iteration 1 can be seen in table 7.

Then the result of the average value of each cluster is used as the new centroid in the next iteration. The average value can be seen in table 8.

After getting the new centroid value, the next iteration is carried out until the centroid value for each cluster does not change. In this study, the calculation process stopped at the 5th iteration. With the value of the centroid of the 4th iteration and the 5th iteration does not change. The results of iteration 5 calculations can be seen in table 9 below.

Table 9 is the final stage of the iteration, the centroid is in accordance with each cluster. Cluster 1 has 72 members, cluster 2 has 456 members, cluster 3 has 74 members, cluster 4 has 463 members, cluster 5 has 87 members and cluster 6 has 1,244 members. The closest distance and cluster members are also listed in the table. From the results of the last iteration, cluster members are formed with the average value as shown in table 10.

Table 7. Iteration Calculation Results 1

Student	C1	C2	C3	C4	C5	C6	Closest distance
Student 1	5	5	2.8284	3	1.4142	4.5826	1.4142
Student 2	6.4031	1	2.8284	5	4.2426	4.5826	1
Student 3	5.8310	2	2.2361	4.2426	3.3166	4.2426	2
Student 4	5.0990	4	2.2361	3.1623	1.7321	4.2426	1.7321
Student 5	5.3852	3	2	3.6056	2.4495	4.1231	2
Student 6	5.0000	5.0000	2.8284	3.0000	1.4142	4.5826	1.4142
Student 7	6.4031	1	2.8284	5	4.2426	4.5826	1
Student 8	5.3852	3	2	3.6056	2.4495	4.1231	2
Student 9	5	5	2.8284	3	1.4142	4.5826	1.4142
...		...	...	...			...
Student 2396	5.1962	4.1231	2.4495	3.3166	1.4142	4.1231	1.4142

Table 8.Iteration Average Value 1

Cluster	V1	V2	V3
C1	3	7.555556	1.833333
C2	7.984018265	4.648402	1.883562
C3	7.738609113	5.997602	1.992806
C4	5.588235294	7.882353	1.882353
C5	7.972383721	7.656977	1.859012
C6	6.053097345	6.283186	1.460177

Table 9.Iteration Calculation Results 5

Student	C1	C2	C3	C4	C5	C6	Closest distance	Cluster
Student	4.6140	3.3521	2.0034	2.2957	0.3805	4.0453	0.3805	C5
Student 2	5.4966	0.6606	2.0034	4.1888	3.6506	2.8555	0.6606	C2
Student 3	4.9982	0.3693	1.0068	3.3878	2.6517	2.6845	0.3693	C2
Student 4	4.5299	2.3529	1.0068	2.3107	0.6636	3.3633	0.6636	C5
Student 5	4.6638	1.3551	0.1169	2.7218	1.6541	2.8738	0.1169	C3
Student 6	4.6140	3.3521	2.0034	2.2957	0.3805	4.0453	0.3805	C5
Student 7	5.4966	0.6606	2.0034	4.1888	3.6506	2.8555	0.6606	C2
Student 8	4.6638	1.3551	0.1169	2.7218	1.6541	2.8738	0.1169	C3
Student 9	4.6140	3.3521	2.0034	2.2957	0.3805	4.0453	0.3805	C5
...				...	...	...	...	...
Student 2396	4.5764	2.5109	1.3384	2.4483	1.0786	3.5013	1.0786	C5

Table 10. The average value of the formed clusters

	V1	V2	V3
Cluster 1	5.315789	4.973684	1.973684
Cluster 2	5.758621	7.534483	1.827586
Cluster 3	3.480769	7.115385	1.711538
Cluster 4	7.963768	6.000000	1.888889
Cluster 5	7.979545	4.650000	1.884091
Cluster 6	7.960545	7.647776	1.861549

Table 10 shows that the average value of cluster 1 is the value of the sub-district, namely 5.315789, the value of parents' work is 4.973684, the value of the study program is 1, 973684 and so on until cluster 6 as shown in table 10..

From the clustering results obtained, the average distance of each jth object is calculated with all objects in the same cluster. Then find the minimum value of the average distance of each object to j-j with all objects in different clusters. Next, look for the SC value of each jth data, so that the coefficient value is the same as the number of datasets. The results of the SC calculation are added up and then divided by the amount of data used. The results of calculating the SC value can be seen in Table 11.

Table 11. SC Calculation Results

Cluster	S(i)
Cluster 1	-0.65814
Cluster 2	0.783079
Cluster 3	-0.67237
Cluster 4	0.799798
Cluster 5	-0.93773
Cluster 6	0.824969

Testing of the number of clusters using 2396 test data, where to find out the best cluster based on the results of the SC value. The cluster values to be tested are k values 1 to 6. In the results of the testing process, the best cluster quality is obtained in cluster 2 with a value of 0.783079, cluster 4 with a value of 0.799798, and cluster 6 with a value of 0.824969 which means that it is included in strong structure category. Cluster 1, cluster 3, and cluster 5 get Si value <= 0.25 which means no structure.

c. Simple Additive Weighting

After obtaining the final centroid of K-means calculation, the next step is calculated using the SAW method to determine the ranking of each cluster member. There are 6 clusters formed from the K-means calculation. The first step is to normalize the data, the value of all members in each attribute is calculated by equation (3). The value included in the benefit attribute is calculated as the max value, and includes the cost attribute which is calculated as the min value. In calculating the member value for each attribute using equation (4).

$$\begin{bmatrix} 5.3 & 5 & 2 \\ 5.8 & 7.5 & 1.8 \\ 3.5 & 7.1 & 1.7 \\ 8 & 6 & 1.9 \\ 8 & 4.7 & 1.9 \\ 8 & 7.6 & 1.9 \end{bmatrix} \tag{4}$$

In solving the value of the attribute, namely by using the SAW method. The first step is to normalize the X matrix.

Normalization V1 :

$$r12 = \frac{8}{\max(5.2,\ 5.8,\ 3.5,\ 8,\ 8,\ 8)} = \frac{8}{8} = 1$$

$$r12 = \frac{8}{\max(5.2,\ 5.8,\ 3.5,\ 8,\ 8,\ 8)} = \frac{8}{8} = 1$$

$$r12 = \frac{8}{\max(5.2,\ 5.8,\ 3.5,\ 8,\ 8,\ 8)} = \frac{8}{8} = 1$$

$$r12 = \frac{8}{\max(5.2,\ 5.8,\ 3.5,\ 8,\ 8,\ 8)} = \frac{8}{8} = 1$$

$$r12 = \frac{8}{\max(5.2,\ 5.8,\ 3.5,\ 8,\ 8,\ 8)} = \frac{8}{8} = 1$$

$$r12 = \frac{8}{\max(5.2,\ 5.8,\ 3.5,\ 8,\ 8,\ 8)} = \frac{8}{8} = 1$$

Normalization V2 :

$$r14 = \frac{7.5}{\max(5,\ 7.5,\ 7.1,\ 6,\ 4.7,\ 7.6)} = \frac{7.5}{7.6} = 0.9852861$$

$$r14 = \frac{7.5}{\max(5,\ 7.5,\ 7.1,\ 6,\ 4.7,\ 7.6)} = \frac{7.5}{7.6} = 0.9852861$$

$$r18 = \frac{7.6}{\max(5,\ 7.5,\ 7.1,\ 6,\ 4.7,\ 7.6)} = \frac{7.6}{7.6} = 1$$

$$r18 = \frac{7.6}{\max(5,\ 7.5,\ 7.1,\ 6,\ 4.7,\ 7.6)} = \frac{7.6}{7.6} = 1$$

$$r18 = \frac{7.6}{\max(5,\ 7.5,\ 7.1,\ 6,\ 4.7,\ 7.6)} = \frac{7.6}{7.6} = 1$$

$$r18 = \frac{7.6}{\max(5,\ 7.5,\ 7.1,\ 6,\ 4.7,\ 7.6)} = \frac{7.6}{7.6} = 1$$

Normalization V3 :

$$r20 = \frac{1.8}{\min(2,\ 1.8,\ 1.7,\ 1.9,\ 1.9,\ 1.9)} = \frac{1.8}{1.7} = 0.936502$$

$$r20 = \frac{1.8}{\min(2,\ 1.8,\ 1.7,\ 1.9,\ 1.9,\ 1.9)} = \frac{1.8}{1.7} = 0.936502$$

$$r24 = \frac{1.9}{\min(2,\ 1.8,\ 1.7,\ 1.9,\ 1.9,\ 1.9)} = \frac{1.9}{1.7} = 0.919416$$

$$r24 = \frac{1.9}{\min(2,\ 1.8,\ 1.7,\ 1.9,\ 1.9,\ 1.9)} = \frac{1.9}{1.7} = 0.919416$$

$$r24 = \frac{1.9}{\min(2,\ 1.8,\ 1.7,\ 1.9,\ 1.9,\ 1.9)} = \frac{1.9}{1.7} = 0.919416$$

$$r24 = \frac{1.9}{\min(2,\ 1.8,\ 1.7,\ 1.9,\ 1.9,\ 1.9)} = \frac{1.9}{1.7} = 0.919416$$

From the results of normalization, the matrix is obtained as shown in table 12.

Table 12. Normalization Results

	V1	V2	V3
cluster 1	0.666177	0.6503439	0.867179
cluster 2	0.721673	0.9851861	0.936502
cluster 3	0.436211	0.9303861	1
cluster 4	0.998023	0.7845418	0.906109
cluster 5	1.000000	0.6080199	0.908416
cluster 6	0.997619	1	0.919416

The next step is the ranking process using the criteria weights that have been given by the decision maker using equation (3). The ranking calculations for alternatives in criteria 1 to criteria 6 are as follows:

Cluster 1 = (0.45* 0.666177) + (0.35* 0.650344) + (0.20* 0.867179)
= 0.700836

Cluster 2 = (0.45* 0.721673) + (0.35* 0.985186) + (0.20* 0.936502)
= 0.856868

Cluster 3 = (0.45* 0.436211) + (0.35* 0.930386) + (0.20*1)
= 0.721930

Cluster 4 = (0.45* 0.998023) + (0.35* 0.784542) + (0.20* 0.906109)
= 0.904922

Cluster 5 = (0.45* 1) + (0.35* 0.608020) + (0.20* 0.908416)
= 0.844490

Cluster 6 = (0.45* 0.997619) + (0.35* 1) + (0.20* 0.919416)
= 0.982812

The results of the alternative ranking calculations that have been sorted can be seen in table 13.

Table 13. Alternative Ranking Calculation Results

No	Cluster	Nilai
1	*Cluster* 6	0.982812
2	*Cluster* 4	0.904922
3	*Cluster* 2	0.856868
4	*Cluster* 5	0.844490
5	*Cluster* 3	0.721930
6	*Cluster* 1	0.700836

The results of the ranking in table 13 show that the highest value is in cluster 6 with a value of 0.982812, then cluster 4 with a value of 0.904922, cluster 2 gets a value of 0.856868, cluster 5 gets a value of 0.844490, cluster 3 gets a value of 0.721930, and cluster 1 gets a final value of 0.700836.

d. Decision Recommendation

Referring to the decision tree that has been determined in table 3, the results of the ranking using the SAW method are used as a recommendation for decision making. The decision tree recommendations is presented in table 14.

Table 14 shows that the first rank is cluster 6 with a value of v1 which is 5.3, the value of v2 is 5, and the value of v3 is 2.The promotion recommended for this cluster is the promotion strategy P1: Social Media, P2: TV-Radio-Newspaper Advertising. Mass media, P3: Brochure distribution, P5: Education Exhibition, P7: Scholarships, and P8: Promotion through invitation. The recommended location for promoting new students in this cluster is the location provided that the distance from the tertiary institution is between 301 and 400 km. Besides that, the average job condition of the parents of students at this location is as laborers.

Table 14. Recommended Decisions

Cluster	V1	V2	V3	Promotion stategy
Cluster 6	5.3	5	2	P1,P2,P3,P5,P8,P7
Cluster 4	5.8	7.5	1.8	P1,P2,P3,P5,P8,P7
Cluster 2	3.5	7.1	1.7	P1,P2,P9,P8,P7
Cluster 5	8	6	1.9	P3,P4,P5,P6,P7
Cluster 3	8	4.7	1.9	P3,P4,P5,P6,P7
Cluster 1	8	7.6	1.9	P3,P4,P5,P6,P7

In cluster 4, it is ranked second with a v1 value of 5.8, a v2 value of 7.5, and a v3 value of 1.8. Promotions recommended for this cluster are P1: Social Media, P2: TV-Radio-Newspaper-Mass Media Ads, P3: Brochure Distribution, P5: Education Exhibition, P7: Scholarships, and P8: Invitation Promotion. The recommended location for promoting new students in this cluster is the location provided that the distance from the university is between 201 km to 300 km. And the average condition of the parents' work at this location is as private employees.

In cluster 2, it is ranked third with a v1 value of 3.5, a v2 value of 7.1, and a v3 value of 1.7. Promotions recommended for this cluster are P1: Social Media, P2: TV-Radio-Newspaper-Mass Media Ads, P7: Scholarships, and P8: Invitation Promotion, and P9: Broadcast SMS. The recommended location for promoting new students in this cluster is the location provided that the distance from the college is above 500 km. And the average condition of the parents' work at this location is as private employees.

In cluster 5, it is ranked third with a v1 value of 8, a v2 value of 6, and a v3 value of 1.9. Promotions recommended for this cluster are P3: Distribution of Brochures, P4: Posting Banners, Pamphlets, Banners, P5: Education Exhibition, P6: Socialization to schools, and P7: Scholarships. The recommended location for promoting new students in this cluster is the location provided that the distance from the college is between 1 km to 101 km. And the average conditions of work for parents of students at this location are PNS / TNI / POLRI (government servicepeople and the military).

Cluster 3 gets the third rank with a v1 value of 8, a v2 value of 4.7, and a v3 value of 1.9. Promotions recommended for this cluster are P3: Distribution of Brochures, P4: Posting Banners, Pamphlets, Banners, P5: Education Exhibition, P6: Socialization to schools, and P7: Scholarships. The recommended location for promoting new students in this cluster is the location provided that the distance from the college is between 1 km to 101 km. And the condition of the average work of the parents of students at this location is as laborers.

In cluster 1, it is ranked third with a v1 value of 8, a v2 value of 7.6, and a v3 value of 1.9. Promotions recommended for this cluster are P3: Distribution of Brochures, P4: Posting Banners, Pamphlets, Banners, P5:

Education Exhibition, P6: Socialization to schools, and P7: Scholarships. The recommended location for promoting new students in this cluster is the location provided that the distance from the college is between 1 km to 101 km. And the average job conditions of parents of students at this location are farmers and fishermen.

Recommendations are presented in the form of patterns that are useful as the output of this study. The output answers the problems formulated from the start. The decision formulation is obtained by presenting the formed clusters and the promotion strategies listed for each cluster.

This study shows that the combination of the k-means method with SAW can be applied to new student admissions data for the purposes of determining how to promote, both in determining the location or strategy of promotion. Data criteria can be varied to suit the needs and presentation of information.

4. Conclusion

The discussion in the Results section shows that the K-means method has produced 6 clusters. The Silhouette coefficient calculation suggests that cluster 2, cluster 4, and cluster 6 belong to a strong structure. Meanwhile, cluster 1, cluster 3, and cluster 5 go into the no structure category. Cluster ranking using the SAW method produces the following sequence: (1) cluster 6, (2) cluster 4, (3) cluster 2, (4) cluster 5, (5) cluster 3, and (6) cluster 1. The ranking provides six recommendations as to potential locations for conducting new student promotions. The decision for location selection agrees with the promotional strategies as described in the decision tree table.

The combination of the K-means and SAW methods produces a recommendation for promotion locations and the strategy that matches the criteria for the particular promotion locations. The management will hopefully use the generated recommendations to make decisions regarding promotion.

The combination of the two methods yet has drawbacks. There is a similarity in the value of the data criteria between one sub variable and another. It has an effect on the results of grouping using the K-means method resulting in the similarity of the calculated data on the sub-variables.

Reference

[1] R. Budiman and Rudianto, "Penerapan Data Mining Untuk Menentukan Lokasi Promosi Penerimaan Mahasiswa Baru Pada Universitas Banten Jaya (Metode K-Means Clustering)," *ProTekInfo(Pengembangan Ris. dan Obs. Tek. Inform.*, vol. 6, no. 1, p. 6, 2019, doi: 10.30656/protekinfo.v6i1.1691.

[2] E. Asril, F. Wiza, and Y. Yunefri, "Analisis Data Lulusan dengan Data Mining untuk Mendukung Strategi Promosi Universitas Lancang Kuning," *J. Teknol. Inf. Komun. Digit. Zo.*, vol. 6, no. 2, pp. 24–32, 2015.

[3] S. Sagala, *Manajemen Strategi dalam Peningkatan Mutu Pendidikan (Pembuka Ruang Kreatifitas, Inovasi, dan Pemberdayaan Potensi Sekolah dalam Sistem Otonomi Sekolah)*. Bandung: Alfabeta, 2007.

[4] S. Abadi *et al.*, "Application model of k-means clustering: Insights into promotion strategy of vocational high school," *Int. J. Eng. Technol.*, vol. 7, no. 2.27 Special Issue 27, pp. 182–187, 2018, doi: 10.14419/ijet.v7i2.11491.

[5] J. O. Ong, "Implementasi Algotritma K-means clustering untuk menentukan strategi marketing president university," *J. Ilm. Tek. Ind.*, vol. vol.12, no, no. juni, pp. 10–20, 2013.

[6] G. A. Sandag, E. Y. Putra, R. L. Wurangian, and N. B. Tulangow, "Analysis of Strategy for Targeted New Student Using K-Means Algorithm," *2019 1st Int. Conf. Cybern. Intell. Syst. ICORIS 2019*, vol. 1, no. August, pp. 94–99, 2019, doi: 10.1109/ICORIS.2019.8874903.

[7] M. Rusli, S. Arifin, and A. Trisnadoli, "Pengembangan Sistem Pendukung Keputusan untuk Penentuan Lokasi Promosi Penerimaan Mahasiswa Baru," *J. Komput. Terap.*, vol. 3, no. 1, pp. 11–18, 2017.

[8] F. Febriadi, A. Ananda, and W. Nengsih, "Sistem Pendukung Keputusan Penentu Daerah Potensial Promosi Perguruan Tinggi Menggunakan Metode Fuzzy AHP (Studi Kasus: Politeknik Caltex Riau)," *J. aksara Komput. Terap.*, vol. 6, no. no 2, 2017.

[9] S. Sugiarti, D. K. Nahulae, T. E. Panggabean, and M. Sianturi, "Sistem Pendukung Keputusan Penentuan Kebijakan Strategi Promosi Kampus Dengan Metode Weighted Aggregated Sum Product Assesment (WASPAS)," *JURIKOM (Jurnal Ris. Komputer)*, vol. 5, no. 2, pp. 103–108, 2018.

[10] S. H. Zanakis, A. Solomon, N. Wishart, and S. Dublish, "Multi-attribute decision making: A simulation comparison of select methods," *Eur. J. Oper. Res.*, vol. 107, no. 3, pp. 507–529, 1998, doi: 10.1016/S0377-2217(97)00147-1.

[11] Sunarti, "Perbandingan Metode TOPSIS dan SAW Untuk Pemilihan Rumah Tinggal," *J. Inf. Syst.*, vol. 3, no. 1, pp. 69–79, 2018.

[12] A. P. Wicaksono, A. Syukur, and Suprapedi, "Komparasi Simple Additive Weighting Dan Analytical Hierarchy Process Terhadap Penentuan Pengelompokan Desa," *J. Teknol. Inf.*, vol. 15, pp. 28–44, 2019.

[13] Muh. Erdiansyah T, B. Pramono, J. Nangi, J. T. Informatika, F. Teknik, and U. H. Oleo, "Perbandingan Metode Fuzzy Analytical Hierarchy Process Dengan Metode Fuzzy Simple Additive Weighting Dalam Menentukan Status Karyawan," *SemanTIK*, vol. 5, no. 2, pp. 231–236, 2019.

[14] D. Korsemov and D. Borissova, "AMO - Advanced Modeling and Optimization, Volume 19, Number 1," *AMO – Adv. Model. Optim.*, vol. 20, no. 1, pp. 101–112, 2018.

[15] K. Arai and A. Ridho Barakbah, "Hierarchical K-means: an algorithm for centroids initialization for K-means," *Rep. Fac. Sci. Engrg. Reports Fac. Sci. Eng.*, vol. 36, no. 1, pp. 25–31, 2007.

[16] A. K. Jain, "Data clustering: 50 years beyond K-means," *Pattern Recognit. Lett.*, vol. 31, no. 8, pp. 651–666, 2010, doi: 10.1016/j.patrec.2009.09.011.

[17] T. Velmurugan and T. Santhanam, "A comparative analysis between K-Medoids and fuzzy C-Means clustering algorithms for statistically distributed data points," *J. Theor. Appl. Inf. Technol.*, vol. 27, no. 1, pp. 19–30, 2011.

[18] X. Wu *et al.*, *Top 10 algorithms in data mining*, vol. 14, no. 1. 2008.

[19] D. J. T. et Al, "Peraturan Presiden Republik Indonesia Nomor 65 Tahun 2011 Tentang Percepatan Pembangunan Provinsi Papua Dan Provinsi Papua Barat Dengan." 2011.

[20] BPPD, "Laporan Akuntabilitas Kinerja Instansi Pemerintah," 2007.

[21] Muhammad Azkaenza, W. J. Pratama, H. Peranginangin, and M. S. Nova, "Laporan Perekonomian Provinsi Papua," vol. 5, p. 98, 2019.

[22] S. Nawrin, M. Rahatur, and S. Akhter, "Exploreing K-Means with Internal Validity Indexes for Data Clustering in Traffic Management System," *Int. J. Adv. Comput. Sci. Appl.*, vol. 8, no. 3, 2017, doi: 10.14569/ijacsa.2017.080337.

[23] D. A. Kandeil, A. A. Saad, and S. M. Youssef, "A two-phase clustering analysis for B2B customer segmentation," *Proc. - 2014 Int. Conf. Intell. Netw. Collab. Syst. IEEE INCoS 2014*, pp. 221–228, 2014, doi: 10.1109/INCoS.2014.49.

[24] P. Rousseeuw, "Rousseeuw et al 1986.pdf," *Journal of Computational and Applied Mathematics*. pp. 53–65, 1986.

[25] J. E. Gentle, L. Kaufman, and P. J. Rousseuw, "Finding Groups in Data: An Introduction to Cluster Analysis.," *Biometrics*, vol. 47, no. 2, p. 788, 1991, doi: 10.2307/2532178.

[26] A. Pranolo and S. M. Widyastuti, "Simple additive weighting method on intelligent agent for urban forest health monitoring," *Proc. - 2014 Int. Conf. Comput. Control. Informatics Its Appl. "New Challenges Oppor. Big Data", IC3INA 2014*, pp. 132–135, 2014, doi: 10.1109/IC3INA.2014.7042614.

[27] J. Kittur *et al.*, "Evaluating Optimal Generation using different Multi– Criteria Decision Making Methods," *Int. Conf. Circuit, Power Comput. Technol. [ICCPCT] Eval.*, 2015.

[28] H. G. Shakouri, M. Nabaee, and S. Aliakbarisani, "A quantitative discussion on the assessment of power supply technologies: DEA (data envelopment analysis) and SAW (simple additive weighting) as complementary methods for the 'Grammar,'" *Energy*, vol. 64, pp. 640–647, 2014, doi: 10.1016/j.energy.2013.10.022.

[29] R. E. Setyani and R. Saputra, "Flood-prone Areas Mapping at Semarang City by Using Simple Additive Weighting Method," *Procedia - Soc. Behav. Sci.*, vol. 227, no. November 2015, pp. 378–386, 2016, doi: 10.1016/j.sbspro.2016.06.089.

[30] A. Faizin and E. Mulyanto, "Penerapan Metode Simple Additive Weighting (SAW) Untuk Seleksi Tenaga Kerja Baru Bagian Produksi (Studi Kasus Pada PT . Jesi Jason Surja Wibowo)," *Univ. Dian Nuswantoro Semarang*, pp. 1–9, 2015.