# Decision Support System for the Success of Education Program at Secondary School Level using Combination of K-Medoids Clustering and TOPSIS

Devi Tri Widyanti [1*], Anggyi Trisnawan Putra [1], Much Azis Muslim [1]

[1]Department Of Computer Science, Faculty of Mathematics and Natural Sciences, Universitas Negeri Semarang, Semarang, Indonesia
*Corresponding author: widyantidevi@gmail.com

ARTICLE INFO

**Article history**
Received 14 February 2020
Revised 19 March 2020
Accepted 3 April 2020

**Keywords**
K-Medoids Clustering
TOPSIS
Decision Support System
Educational Program

ABSTRACT

The success of education programs is one of the concepts of educational equity that aims to educate a nation's life. However, the condition of education in Indonesia is not evenly distributed, and this can be seen from the availability and affordability of education services in each province. This research applies a decision support system to determine two categories: provinces that have not achieved a success rate of education program using K-Medoids Clustering and TOPSIS. The K-Medoids Clustering was used to overcome the outliers in the data. The TOPSIS is used to provide decision making based on the best alternative concept. The best alternative concept in TOPSIS has the closest distance from the positive ideal solution and the farthest distance from the negative ideal solution. The number of clusters formed as many as five clusters. The iteration needed to cluster provinces using K-Medoids Clustering is 817 iterations. The third cluster has the largest variable average value and smallest standard deviation value. So, the third cluster shows the best cluster quality. Determination of provincial members into two categories is partially/fully refers to members of best cluster quality and TOPSIS preference value.



## 1 Introduction

Data mining is a technique to dig valuable information buried or hidden in a data collection that is so big to find an interesting pattern that was previously unknown (Sugiharti, Firmansyah, & Devi, 2017). Clustering is one of the data mining methods. One of the main requirements of any clustering algorithm is a good similarity measure to know the distance between the objects in order to group them together (Harikumar & Surya, 2015). Partitioning based clustering is grouping data from one large group then divided into smaller groups (Defiyanti, Jajuli, & Rohmawati, 2017). K-Medoids Clustering is part of the partitioning clustering method. The K-Medoids Clustering algorithm uses the middle value (medoids) in determining the centroid. The centroid is used to calculate the distance of a data object on the centroid (Nurzahputra, Muslim, & Khusniati, 2017). K-Medoids Clustering iteratively optimizes the medoids object by swapping between medoids object and non-medoids object, then evaluating the swap's cost by repartitioning the data (Rangel, Hendrix, Agrawal, Liao, & Choudhary, 2016).

Decision making is done by a systematic approach to the problem through the process of collecting data into information and factors that need to be considered in decision making (Kurniasih, 2013). The decision support system can be interpreted as a model-based system consisting of procedures in processing the data. The results of the data processing are used to assist managers in making the decision (Sudarma, Sudana, & Cahya, 2015). There are several models that can be used to build a Decision Support System (DSS), one of which is Technique for Order Preference by Similarity to

Ideal Solution (TOPSIS). TOPSIS is based on the concept where the best alternative does not only have the closest distance from the positive ideal solution but also has the farthest distance from the negative ideal solution.

The basis of strategic development is a factor of education. The success of education programs is one of the concepts of educational equity that aims to educate the life of a nation as stated in Law Number 20 of 2003 article 4 paragraph (1) that every Indonesian citizen has the right to obtain education by not carrying out discriminatory actions by upholding human rights, religious values, cultural values, and national plurality. The success of the education program refers to several education indicators according to the education missions. The data analysis technique carried out by the Ministry of Education and Culture through the Education and Culture Statistics Data Center is a descriptive analysis using ideal standards. The ideal standard in question is to assess each education mission and divided it into five categories. Grouping categories, the success of education program is a series of selection in determining the provinces that have not achieved the success rate of the education program.

The education circumstances in Indonesia are not evenly distributed at all school levels, and this can be seen from the availability and affordability of education services in each province. The development of methods or techniques in processing data is very important. This research applies the field of computer science, including data mining methods using the K-Medoids Clustering algorithm and decision support system methods using the TOPSIS algorithm for making the decision in grouping provinces into two categories, namely provinces that have and have not achieved a success rate of the education program at secondary school level in Indonesia.

## 2 Methods

The method used as a reference in this research is K-Medoids Clustering, interpretation and evaluation of cluster results, and TOPSIS.

### 2.1 K-Medoids Clustering

Data processing begins with analyzing clusters using the K-Medoids Clustering method. The clustering principle is to maximize the similarity between members of the classes and minimize the similarity between clusters (Sugiharti & Muslim, 2016). K-Medoids Clustering is part of the partitioning clustering method where data consisting of n objects is partitioned into k clusters, the number of k ≤ n (Wuryandari, Rusgiyono, & Setyowati, 2016). Cluster formation is based on calculating the proximity between medoids object and non-medoids object using Euclidean Distance calculation (Satoto, Khotimah, & Iswati, 2015). Optimizing the medoids object is done by swapping between medoids object and non-medoids object, then evaluating the cost of the swap by repartitioning the data.

The steps of the K-Medoids Clustering method are as follows.

1. Arbitrarily choose $k$ objects as the initial medoids object ($O_c$).
2. Calculate the similarity between medoids object and non-medoids object using Euclidean Distance. The Euclidean Distance formula is as follows.

$$d(i,k) = \sqrt{\sum_{j=1}^{n} \left| x_{ij} - y_{kj} \right|^2} \tag{1}$$

where as $i$ : non-medoids object
$k$ : medoids object
$j$ : variables
$d(i,k)$ : euclidean distance value
$x_{ij}$ : value of the non-medoid object to $i$ in the variable to $j$
$y_{kj}$ : value of the medoid object to $k$ in the variable to $j$
$n$ : number of variables

3. Assign each non-medoid object to a cluster with the nearest medoids object.
4. Randomly select a non-medoids object to replace the initial medoids object ($O_{random}$).

5. Repeat steps 2 and 3, then calculate the value of absolute error ($E$) before and after the swap process. If $E_{random}$ < $E_c$, then swap $O_c$ with $O_{random}$ but if $E_{random}$ > $O_c$ then there is no swap $O_c$. The absolute error ($E$) formula is as follows.

$$E = \sum_{c=1}^{k} \sum_{i=1}^{n_c} |p_{ic} - O_c| \quad (2)$$

whereas $E$ : the sum of absolute error for all items in the data set
$p_{ic}$ : non-medoid object to $i$ in the cluster to $c$
$O_c$ : the value of the medoids object in the cluster to $c$

6. Repeat step 4 and 5 till the medoids object stabilize their locations.

The flowchart of the K-Medoids Clustering algorithm is shown in Figure 1.

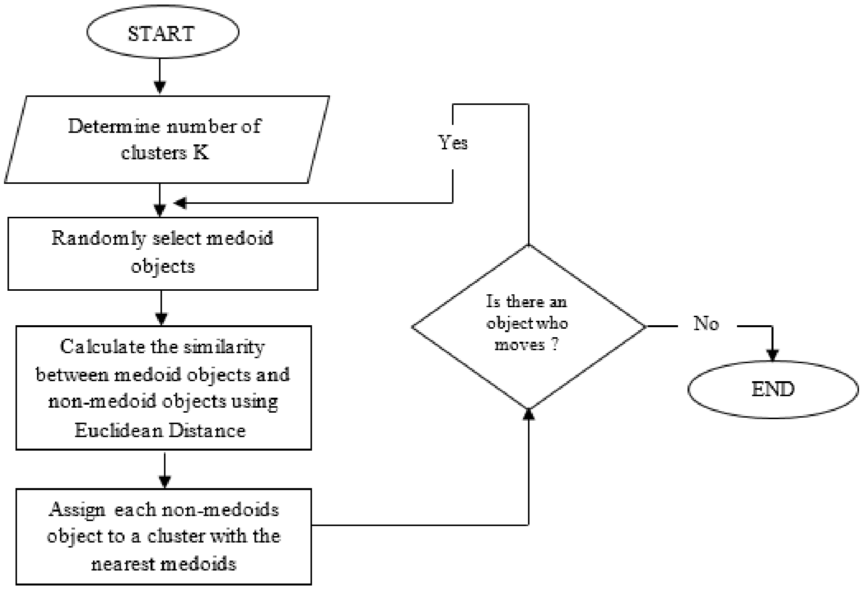


**Figure 1.** Flowchart of the k-medoids clustering algorithm

## 2.2 Interpretation of Cluster Results

Interpretation of cluster results is made after the cluster is formed, which essentially gives a specific name to describe the contents of the cluster (Goreti, Novia, Wahyuningsih, 2016). To interpret clusters and make profiles of all clusters seen from all variables' average value in each cluster formed (Simamora, 2005).

## 2.3 Evaluation of Cluster Results

Standard deviation is used to see the diversity and uniformity of members of each cluster. It can also be referred to as internal validation. Internal validation is used to evaluate the algorithm of clustering data that refers to the assumption that the cluster that is sought must be based on the proximity and unproximity of cluster membership (Hair, Black, Babik, & Anderson, 2010). The smaller the standard deviation, the cluster has good quality, which means a high homogeneity. The standard deviation formula is as follows.

$$s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2 \quad (3)$$

whereas $s$ : standard deviation value
$n$ : number of variables
$x_i$ : variable data value to $i$
$\bar{x}$ : average value of the data variable to $i$

### 2.4 Combination Method of K-Medoids Clustering and TOPSIS

A decision support system takes into account all variables that support decision making to assist, accelerate and simplify the decision-making process (Rahmanda, Arifudin, & Muslim, 2017). TOPSIS is based on a concept where the best alternative has the closest distance from the positive ideal solution and has the farthest distance from the negative ideal solution. A positive ideal solution maximizes the benefit attribute and minimizes the cost attribute, while the ideal negative solution maximizes the cost attribute and minimizes the benefit attribute (Adi, Sugiharti, & Alamsyah, 2018). The steps of the TOPSIS method are as follows.

1. Make a decision matrix according to the problem to be solved, then make a normalized decision matrix using Equation 4.

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

where as $i$ : 1,2,3,...m (row of each decision matrix)
$j$ : 1,2,3,...m (column of each decision matrix)
$r_{ij}$ : the normalized decision matrix
$x_{ij}$ : the original decision matrix

2. Create a normalized weighted decision matrix using Equation 5.

$$y_{ij} = w_i * r_{ij} \tag{5}$$

whereas $y_{ij}$ : the normalized weighted decision matrix
$w_i$ : the weight value of each variable

3. Determine the positive ideal solution ($A^+$) and negative ideal solution ($A^-$) based on the normalized weighted decision matrix by using Equations 6 and 7.

$$A^+ = ({y_1}^+, {y_2}^+, \dots, {y_n}^+) \tag{6}$$

$$A^- = ({y_1}^-, {y_2}^-, \dots, {y_n}^-) \tag{7}$$

$${y_i}^+ = \begin{cases} max\ y_{ij}\ ; if\ j\ is\ a\ benefit\ attribute \\ min\ y_{ij}\ ; if\ j\ is\ a\ cost\ attribute \end{cases} \tag{8}$$

$${y_i}^- = \begin{cases} max\ y_{ij}\ ; if\ j\ is\ a\ cost\ attribute \\ min\ y_{ij}\ ; if\ j\ is\ a\ benefit\ attribute \end{cases} \tag{9}$$

whereas $A^+$ : the positive ideal solution
$A^-$ : the negative ideal solution
${y_i}^+$ : element of the positive ideal solution
${y_i}^-$ : element of the negative ideal solution

4. Determine the distance between the normalized weighted decision matrix with the positive ideal solution and negative ideal solution using Equations 10 and 11.

$${D_i}^+ = \sqrt{\sum_{i=1}^{n} ({y_i}^+ - y_{ij})^2} \tag{10}$$

$${D_i}^- = \sqrt{\sum_{i=1}^{n} (y_{ij} - {y_i}^-)^2} \tag{11}$$

whereas ${D_i}^+$ : distance to a positive ideal solution
${D_i}^-$ : distance to a negative ideal solution

5. Determine the preference value for each alternative by using Equation 12.

$$V_i = \frac{D_i^-}{D_i^- + D_i^+} \tag{12}$$

whereas $V_i$ : preference value for each alternative

The combination of K-Medoids Clustering and the TOPSIS method in this research aims to help make the final decision in grouping provinces into two categories: provinces that have and have not achieved the education program's success rate at the secondary school level in Indonesia. Determination of provincial members into two categories is partially or fully refers to members of best cluster quality and TOPSIS preference value. The flowchart of the combination of K-Medoids Clustering and TOPSIS method is shown in Figure 2.

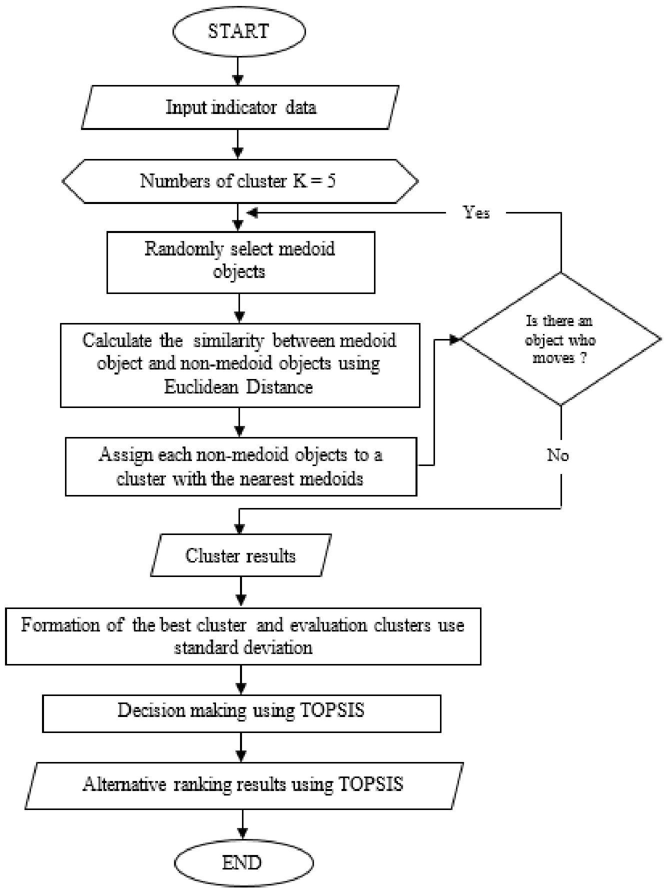


**Figure 2.** Flowchart of the combination of k-medoids clustering and topsis method

## 3 Results and Discussion

The data used in this research is province data in Indonesia as many as 34 provinces as alternatives data and education indicators data as many as nine indicators as variables data. These data were obtained from the Ministry of Education and Culture by the publication of the Education and Culture Statistics Data Center at the secondary school level in 2016. Education indicators used as variables data are Student/School Ratio (R-S/Sek), Student/Class Ratio (R-S/K), Class/Classroom Ratio (R-K/RK), Percentage of School Library (%Perpus), Percentage of Laboratory (%Lab), Gross Participation Rate (APK), Continuing Numbers (SMA/equivalent) (AM), Gender Comparison APK (PG-APK) and Gender Parity Index APK (IPG-APK).

The first step is the K-Medoids Clustering process. The number of clusters used is five clusters because the Education and Culture Statistics Data Center's ideal standard to assess each education mission was divided into five categories. Medoids were selected randomly as the number of clusters. For example, the initial medoids objects included Bengkulu, Bangka Belitung, West Java, Maluku, and North Maluku. Then calculate on each distance between the medoids object with the non-medoids object, the education indicators data uses Euclidean Distance through Equation (1). Then the distance is chosen to be clustered based on the closest distance to the medoids object. Random selection of the non-medoids object to replace the initial medoids object is repeated in the next iteration. Then calculate the Euclidean Distance and cluster again. Next, calculate the value of absolute error ($E$) before and after the swap process through Equation (2). Iteration is continued by repeating the previous step till the medoids object stabilize their locations. The number of iterations needed till the medoids object stabilizes their locations as many as 817 iterations. Clustering results using the K-Medoids clustering method are shown in Table 1.

**Table 3.** Achievement of education program

| Provinces | $V_i$ | Source |
|---|---|---|
| South Sumatera | - | K-Medoid |
| Bengkulu | - | K-Medoid |
| Bangka Belitung Island | - | K-Medoid |
| Central Java | - | K-Medoid |
| North Sulawesi | - | K-Medoid |
| South Sulawesi | - | K-Medoid |
| Yogyakarta | 0,867429 | TOPSIS |
| Gorontalo | 0,847678 | TOPSIS |
| Southeast Sulawesi | 0,828177 | TOPSIS |
| Jakarta | 0,8233 | TOPSIS |

**Table 4.** Unachievement of education program

| Provinces | $V_i$ |
|---|---|
| Papua | 0,399344 |
| Riau Island | 0,512019 |
| West Papua | 0,608033 |
| West Nusa Tenggara | 0,62363 |
| South Kalimantan | 0,632931 |
| Central Kalimantan | 0,636908 |
| Banten | 0,64771 |
| Maluku | 0,666391 |
| North Kalimantan | 0,670917 |
| West Sulawesi | 0,678366 |

## 4 Conclusion

The combination of the K-Medoids Clustering and TOPSIS method aims to group provinces in Indonesia into two categories as the final result of decision making. These categories are provinces that have and have not achieved the success rate of education programs in Indonesia. In this research, based on the application of the K-Medoids Clustering method, the number of iterations needed until the medoids object stabilized their location was 817 iterations. After the cluster is formed, the third cluster has the largest variable average value and has a small standard deviation value. So, it can be concluded that the third cluster shows the best cluster quality. Members of the third cluster are categorized as provinces that have achieved the education program's success rate at the secondary school level. The TOPSIS calculation is applied to determine the provinces that have not achieved the education program's success rate. List of provinces included in these two categories, each of which is ten provinces. If the number of members of the cluster that have achieved the education program's success rate is still less than ten provinces, then take the provinces by obtaining a high preference value from the TOPSIS calculation.

## References

Adi, P. T. K., Sugiharti, E., & Alamsyah. (2018). Comparison Between SAW and TOPSIS Methods in Selection of Broiler Chicken Meat Quality. *Scientific Journal of Informatics, 5*(1), 81-90. doi:10.15294/sji.v5i1.14416

Defiyanti, S., Jajuli, M., & Rohmawati, N. (2017). Optimalisasi K-Medoid dalam Pengklasteran Mahasiswa Pelamar Beasiswa dengan Cubic Clustering Criterion [Optimizing K-Medoid in Clustering Scholarship Applicants with Cubic Clustering Criterion]. *TEKNOSI, 3*(1), 211-218. doi:10.25077/TEKNOSI.v3i1.2017.211-218

Goreti, M., Novia, Y., & Wahyuningsih, S. (2016). Perbandingan Hasil Analisis Cluster dengan menggunakan Metode Single Linkage dan Metode C-Means [Comparison of Cluster Analysis Results using the Single Linkage Method and the C-Means Method]. *EKSPONENSIAL, 7*(1), 9-16. Retrieved from http://jurnal.fmipa.unmul.ac.id/index.php/exponensial/article/view/15

Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2010). *Multivariate Data Analysis: A Global Perspective (7th Ed.)*. Upper Saddle River, NJ: Prentice Hall.

Harikumar, S., & Surya, P. V. (2015). K-Medoid Clustering for Heterogeneous DataSets. *Procedia Computer Science, 70*, 226-237. doi:10.1016/j.procs.2015.10.077

Kurniasih, D. L. (2013). Sistem Pendukung Keputusan Pemilihan Laptop dengan Metode TOPSIS [Laptop Selection Decision Support System with the TOPSIS Method]. *Pelita Informatika Budi Darma, 3*(2), 6-13.

Nurzahputra, A., Muslim, M. A., & Khusniati, M. (2017). Penerapan Algoritma K-Means untuk Clustering Penilaian Dosen berdasarkan Indeks Kepuasan Mahasiswa [Application of the K-Means Algorithm for Clustering of Lecturer Ratings based on the Student Satisfaction Index]. *Techno.Com, 16*(1), 17-24. doi:10.33633/tc.v16i1.1284

Rahmanda, P. O., Arifudin, R., & Muslim, M. A. (2017). Implementation of Analytic Network Process Method on Decision Support System of Determination of Scholarship Recipient at House of Lazis Charity UNNES. *Scientific Journal of Informatics, 4*(2), 199-211. doi:10.15294/sji.v4i2.11852

Rangel, E. M., Hendrix, W., Agrawal, A., Liao, W., & Choudhary, A. (2016). AGORAS: A Fast Algorithm for Estimating Medoids in Large DataSets. *Procedia Computer Science, 80*, 1159-1169. doi:10.1016/j.procs.2016.05.446

Satoto, B. D., Khotimah, B. K., & Iswati. (2015). Pengelompokkan Wilayah Madura berdasar Indikator Pemerataan Pendidikan menggunakan Partition Around Medoids dan Validasi Adjusted Random Index [Madura Region Grouping based on Equitable Education

Indicators using Partition Around Medoids and Adjusted Random Index Validation]. *Journal of Information Systems Engineering and Business Intelligence*, *1*(1), 17-24. doi:10.20473/jisebi.1.1.17-24

Simamora, B. (2005). *Analisis Multivariat Pemasaran [Multivariate Marketing Analysis]*. Jakarta: PT Gramedia Pustaka Utama.

Sudarma, M., Sudana, A. A. K. O., & Cahya, I. (2015). Decision Support System for the Selection of Courses in the Higher Education using the Method of Elimination Et Choix Tranduit La Realite. *International Journal of Electrical and Computer Engineering (IJECE)*, *5*(1), 129-135. doi:10.11591/ijece.v5i1.pp129-135

Sugiharti, E. & Muslim, M. A. (2016). On-Line Clustering of Lecturers Performance of Computer Science Department of Semarang State University using K-Means Algorithm. *Journal of Theoretical and Applied Information Technology,* 83(1), 64-71. Retrieved from https://lib.unnes.ac.id/33054/

Sugiharti, E., Firmansyah, S., & Devi, F. R. (2017). Predictive Evaluation of Performance of Computer Science Students of UNNES using Data Mining based on Naïve Bayes Classifier (NBC) Algorithm. *Journal of Theoretical and Applied Information Technology*, *95*(4), 902-911. Retrieved from https://lib.unnes.ac.id/33089/

Wuryandari, T., Rusgiyono, A., & Setyowati, E. (2016). Pengelompokkan Kabupaten/Kota berdasarkan Komoditas Pertanian menggunakan Metode K-Medoids [District / City Grouping based on Agricultural Commodities using the K-Medoids Method]. *Media Statistika*, *9*(1), 41-49. doi:10.14710/medstat.9.1.41-49