Optimization Of Clustering Algorithm On Decision Support System Of Scholarship Recipients Using Analytical Hierarchy Process Method

R Andriani¹, R F Amanullah², D Ninosari³ and Kusrini⁴

¹,²,³,⁴ Magister of Informatics Engineering, Universitas AMIKOM Yogyakarta Yogyakarta, Indonesia.

E-mail: ria.an@students.amikom.ac.id

Abstract. Education has a very important part in the life of the population, but the cost of education often occurs in the study. Therefore, the government provides support in the form of scholarships for underprivileged or achievers students. The scholarship will be awarded to candidates who compatible with the criteria based on government rules. To assist in decision-making on candidates who are eligible to receive a scholarship, a decision support system is required. This study aims to provide recommendations in determining the students who are eligible to receive scholarships. The K-Means Clustering method is used to classify prospective scholarship recipients into agreed parts. The data that has been grouped, will later be processed into the AHP method, in order to obtain recommendations for prospective scholarship recipients. In achieving these recommendations, there will certainly be errors. To avoid this, the Xie Beni Index method is used, in order to provide optimal results for group division based on existing data, which will then be reprocessed in the AHP method. The results showed that the cluster suitable for the Achievement Index, the number of Semester Credit Unit, Parent’s income and dependents of parents were 4, 2, 3, 2. With Xie-Beni Index value of each 0.00267, 0.0022, 0.0025, 0.00384.

1. Introduction

Education has a very important role in the life of the community, because education is considered capable of improving human living standards [1]. It has also been clarified in Law No.20 / 2003 on education which contains efforts made in a conscious and planned to create an atmosphere of learning and learning process so that learners are actively in developing selfpotential [2].

Therefore, the government provides assistance in the form of scholarships. The scholarship is a form of financial assistance provided to students who aim to support the continuity of education that is being pursued, the scholarship is also the right solution to alleviate the cost of education for students so that more and more people who travel education to college level. There are many
types of scholarships in every college, ranging from scholarships to underprivileged students and scholarships to outstanding students.

This scholarship is made through the selection process based on the criteria determined by each university. The problems that often arise in the selection process determining the students who are eligible to receive scholarship is the data that many registrants in each of the criteria that have been determined and still use the manual system so it takes a very long time to group the data and will hamper the selection process others.

In this study the criteria used are Achievement Index, Semester Credit Unit amount, Parent’s Income and Dependence Of Parents. The purpose of this study is to produce a recommendation of students who are eligible to receive scholarships. After collecting the data, then using the K-means Clustering method to classify the data of prospective scholarship recipients into several sections according to the part that has been entered. the data that has been grouped, will then be processed using the AHP method, in order to obtain recommendations for prospective scholarship recipients. to find out the truth of the clustering process, the Xie-Beni Index method is used, in order to get the appropriate number of group divisions based on the data of existing scholarship recipients, which will then be counted again with the AHP method, so that the recipients receive valid recommendations based on registrant data.

The research using K-Mean clustering has been done by Asep et al in 2017 under the title "Design of Decision Support System of Scholarship Receiver Using K-Means Clustering Method" This research aims to design the scholarship decision support system using clustering method [3] . Similar research was also conducted by Josi and Wirda in 2016 entitled "Application of K-Means Cluster Analysis Method In Decision Support System for Selection of Concentration for International Class Students of STMIK AMIKOM" This study aims to create a decision support system for concentration selection by using clustering method [4]. Similar research was also conducted by Ahlihi et al in 2014 under the title "Decision Support System for Determination of Location Site Using K-Means Clustering and TOPSIS" This research aims to make decision support system for determining tourist location by using clustering [5]. Research [6] aims to optimize the grouping of previously processed data using FCM Algorithm (Fuzzy C-Means). So as to produce the value of the objective function that is brought threshold and has the lowest Xie Beni Index value.

Based on the problems that emerged in this study will be made a decision support system of acceptance of scholarships using AHP method by optimizing the clustering algorithm that can help the University of AMIKOM in determining the students who are eligible to get scholarships based on predetermined criteria, the algorithm in this study can also be developed for further research.
2. Methodology

This study began with data collection, which was then grouped using the K-Means Clustering method. The grouping is done based on the number of groups that have been inputted previously, after the grouping is done, the data will be used as input in the AHP method. The results of the AHP method are recommendations for prospective scholarship recipients. To provide optimal results, a testing process is performed using the Xie Beni Index method, this method plays a role in checking the number of data groupings previously carried out by the K-Means Clustering method.

2.1. Data Collection

In this research, the criteria used are: GPA, Number of credits, Semester, Non-Academic Achievement, Income and Parent Dependence. While the dataset used is the data of prospective scholarship recipients at the University of AMIKOM Yogyakarta in 2018 with a sample of 10 students.

2.2. K-Means Clustering

From the data that has been collected in the previous process, then done grouping data based on predetermined criteria. In the grouping there is a calculation process to determine the closeness between each data. The calculation process consists of determining the center of the cluster expressed by the equation:

\[ c_{k,j} = \frac{x_{1j} + x_{2j} + \ldots + x_{nj}}{n} \]  

(1)

With \( c_{k,j} \) being the center of the k-cluster in the variables to \( j = 1, 2, \ldots, p \) and \( n \) are the number of data in the k-cluster.

Calculation of Euclidean distance expressed by equation:

\[ d(x_i, c_{kj}) = \left[ \sum_{i=1}^{p} (x_i - c_{kj})^2 \right]^{1/2} \]  

(2)

The calculation of the objective function is expressed by the equation:

\[ J = \sum_{i=1}^{n} \sum_{j=1}^{j} a_{ij} d(x_i, c_{kj})^2 \]  

(3)

With \( a_{ij} \) is the value of the membership of point \( x_i \) to the \( c_{kj} \) cluster center, whereas \( d(x_i, c_{kj}) \) is the Euclidean distance of the data point \( x_i \) with the \( c_{kj} \) cluster center.

2.3. Testing

To avoid errors in data grouping, the Xie-Beni Index Method is used. This method plays a role in minimizing errors by grouping data based on the total data of prospective scholarship recipients. Which is the result of this result which will be processed in AHP calculation. The calculation of Xie-Beni index is expressed by equation:
\[ S = \frac{\sum_{i=1}^{c} \sum_{j=1}^{n} (u_{ij})^m ||x_j - v_i||^2}{n \sum_{i,j} ||v_i - v_j||^2} \] (4)

Where \( S \) represents the Xie Beni Index, \( c \) is the number of clusters/groups, \( n \) denotes the number of research objects, \( u_{ij} \) represents the membership value of the j object with center group i (value = 1), \( m \) denotes the fuzzifier value (value = 2), \( ||x_j - v_i|| \) declare the distance of Euclidean data point \( x_j \) with center group \( v_i \), while \( ||v_i - v_j|| \) Euclidean distance between the center of the group. The smaller the value of \( S \), the clustering is more accurate.

### 2.4. Conversion of Student Values

At this stage, it is the stage of converting the original values into values based on clusters according to clusters at the K-Means Clustering and Testing Stages (Xie-Beni Index Approach).

### 2.5. AHP Calculation

The converted data will then be used as input in the AHP calculation. In this stage, will be generated recommendations to candidates who are eligible for a scholarship. But in getting these results, there are several processes that need to be done. These processes include: pairwise comparison matrix process, expressed by the equation:

\[ C_i = \sum_j M_{ij} S_j S_i \] (5)

Where \( M_{ij} \) is the value of alternative choice i from alternative j on given criteria. The process of normalization expressed by the equation:

\[ r_{ij} = \frac{x_{ij}}{\sum_{i=1}^{n} X_{i1}} \] (6)

With \( x_{ij} \) representing the value of row i column j, whereas \( X_{i1} \) represents the sum of values from the ith row of column j. Looking for priority weight and lambda value; looking for value of Consistency Index (CI) expressed by the equation:

\[ CI = \frac{\lambda - n}{n - 1} \] (7)

Looking for Ratio Consistency value expressed by the equation:

\[ CR = \frac{CI}{RI} \] (8)

With the value of \( \lambda \) is the average of the consistency value, then for the RI value is based on the amount of data used. For the determination of RI value will be presented in Table 1.
Table 1. The Value Of RI

| N | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| RI | 0.00 | 0.00 | 0.58 | 0.90 | 1.12 | 1.24 |

If the CR value obtained from (8) where there is a division between CI values, which is obtained from (7) and the RI value from Table 1 is less than 10%, then the CR is not consistent, and if the CR value is more than 10% then it will be input into (9)

\[
CR = \sum_{j=1} W_j r_{ij} \tag{9}
\]

3. Result and Discussion

In order to optimize the clustering algorithm on the scholarship decision support system with Analytical Hierarchy Process (AHP) method there are several steps. The steps are shown in Figure 1.

![Figure 1. The Research flow](image)

In accordance with the flow presented in Figure 1. This research was carried out with two processes, where the first process is the calculation process without optimizing the number of groupings, namely the grouping of data using K-Means Clustering and then used as input in the AHP method. While the second process is testing, where the process is used the Xie Beni Index method to provide optimal results in data grouping, which is then also used as input in the AHP method.
3.1. K-Means Clustering

The distribution of the number of clusters conducted randomly or randomly on each criteria, in accordance with Table 2.

| Criteria                  | Number Of Cluster |
|---------------------------|-------------------|
| Achievement Index         | 3                 |
| Number Of Semester credits| 2                 |
| Parental income           | 4                 |
| Parent Dependency         | 2                 |

In this research, the Semester and Non-Academic Achievement criteria are not clustered because the amount has been determined based on government regulations. From the distribution of the number of clusters is then carried out the conversion value, so that the value of students will be seen in Table 3.

| Mhs | C1 | C2 | C3 | C4 | C5 | C6 |
|-----|----|----|----|----|----|----|
| A01 | 2  | 1  | 1  | 0  | 4  | 1  |
| A02 | 3  | 2  | 3  | 2  | 1  | 2  |
| A03 | 1  | 2  | 3  | 0  | 2  | 1  |
| A04 | 1  | 1  | 1  | 0  | 2  | 1  |
| A05 | 3  | 1  | 1  | 1  | 1  | 1  |
| A06 | 1  | 2  | 3  | 0  | 4  | 2  |
| A07 | 1  | 1  | 1  | 0  | 3  | 1  |
| A08 | 1  | 2  | 3  | 0  | 1  | 1  |
| A09 | 2  | 1  | 1  | 2  | 1  | 1  |
| A10 | 3  | 1  | 1  | 0  | 2  | 2  |

The Achievement Index (C1), Number of credits (C2), Semester (C3), Non-Academic Achievement (C4), Parent Earnings (C5) and Parent Dependence (C6).

3.2. Approach of Xie-Beni Index

Based on the data in Table 3, checking was done using the Xie Beni Index. Which is done by checking using (4). The results of the Xie-Beni Index approach are presented in Table 4.
Table 4. Approach of Xie-Beni Index

| Lots Of Data     | Criteria                  | Number Of Cluster | Result   |
|------------------|---------------------------|-------------------|----------|
|                  | Achievement Index (C1)    | 2                 | 0.00285  |
|                  |                           | 3                 | 0.00272  |
|                  |                           | 4                 | 0.00267  |
|                  |                           | 5                 | 0.00625  |
|                  |                           | 2                 | 0.0022   |
|                  | Number of credits (C2)    | 3                 | 0.0025   |
|                  |                           | 4                 | 0.0122   |
|                  |                           | 5                 | 0.01     |
|                  | Parental income (C5)      | 2                 | 0.004    |
|                  |                           | 3                 | 0.0025   |
|                  |                           | 4                 | 0.00571  |
|                  |                           | 5                 | 0.16667  |
|                  | Parent Dependency (C6)    | 2                 | 0.004    |
|                  |                           | 3                 | 0.00428  |
|                  |                           | 4                 | 0.004    |
|                  |                           | 5                 | 0.00444  |

After checking using the Xie Beni Index of the 10 existing scholarship recipient data, then the scores are obtained as in table 4. The next process is to convert the value of the prospective scholarship recipients into groups or groups that have been known from optimization using the Xie Beni Index. The conversion results can then be seen in Table 5.

Table 5. Value Conversion Using the Xie-Beni Index Approach

| Mhs   | C1 | C2 | C3 | C4 | C5 | C6 |
|-------|----|----|----|----|----|----|
| A01   | 2  | 1  | 1  | 0  | 4  | 1  |
| A02   | 3  | 2  | 3  | 2  | 1  | 2  |
| A03   | 1  | 2  | 3  | 0  | 2  | 1  |
| A04   | 1  | 1  | 1  | 0  | 2  | 1  |
| A05   | 3  | 1  | 1  | 1  | 1  | 1  |
| A06   | 1  | 2  | 3  | 0  | 4  | 2  |
| A07   | 1  | 1  | 1  | 0  | 3  | 1  |
| A08   | 1  | 2  | 3  | 0  | 1  | 1  |
| A09   | 2  | 1  | 1  | 2  | 1  | 1  |
| A10   | 3  | 1  | 1  | 0  | 2  | 2  |

3.3. Analytical Hierarchy Process (AHP)
Prior to the ranking, first create a comparison matrix for the criteria, the following matrix of comparison criteria used will be presented in Table 6.
The results of the recommendations with AHP and K-Means Clustering methods are presented in Table 7.

Table 7. AHP Method with K-Means Clustering

| Kode | Nilai   | Rank |
|------|---------|------|
| A02  | 0,14087 | 1    |
| A09  | 0,13563 | 2    |
| A05  | 0,11019 | 3    |
| A06  | 0,09418 | 4    |
| A01  | 0,09359 | 5    |
| A03  | 0,08881 | 6    |
| A07  | 0,08677 | 7    |
| A08  | 0,08432 | 8    |
| A10  | 0,08331 | 9    |
| A04  | 0,08228 | 10   |

From the results of the recommendations contained in Table 7, then optimization of the number of clusters on prospective scholarship recipients using the AHP method is contained in Table 8.

Table 8. Cluster Optimization on AHP Method

| Kode | Nilai   | Rank |
|------|---------|------|
| A02  | 0,14563 | 1    |
| A09  | 0,13382 | 2    |
| A05  | 0,10781 | 3    |
| A06  | 0,09482 | 4    |
| A03  | 0,09347 | 5    |
| A08  | 0,09024 | 6    |
| A01  | 0,08823 | 7    |
| A07  | 0,08475 | 8    |
| A10  | 0,0814  | 9    |
| A04  | 0,07978 | 10   |

From the use of the K-Means Clustering, Xie Beni and AHP methods, two results were obtained where the first results (K-Means Clustering + AHP) are shown in Table 7. And the second result that has been done by grouping optimization using the Xie Beni and AHP indexes is shown by Table 8. From the two results, different data are obtained for the locations of A03, A07, A08.
and A01. Thus the solution is proven, by applying the Xie Beni Index can reduce the risk of errors.

4. Conclusion

This is an example. This is an example. This is an example. This is an example. Please make sure you read the author guidelines to clarify how the article is written. This template is just an example.

From the results of research conducted, the conclusions obtained are as follows: (1) The results of this study in the form of recommendations as a reference of the University in providing scholarships both scholarship achievement and scholarships for the poor; (2) Based on the calculation of Xie-Beni validity index, the optimum number of clusters for achievement index is 4, the number of credits 2 clusters, the income of the parents 3 clusters, the dependents of 2 people cluster; (3) Based on the tests conducted on the proximity of each cluster nodes based on the criteria used got results that the Achievement Index 4 there are 4 clusters with a value of 0.00267, Number of Cluster 2 with value 0.0022, Parents income 3 with value 0.0025 and dependents parents 2 clusters with a value of 0.00384.

5. References

[1] Sze S M 1969 Physics of Semiconductor Devices (New York: Wiley–Interscience)
[2] Dorman L I 1975 Variations of Galactic Cosmic Rays (Moscow: Moscow State University Press) p 103
[3] Caplar R and Kulisic P 1973 Proc. Int. Conf. on Nuclear Physics (Munich) vol 1 (Amsterdam: North-Holland/American Elsevier) p 517
[4] Szytula A and Leciejewicz J 1989 Handbook on the Physics and Chemistry of Rare Earths vol 12, ed K A Gschneidner Jr and L Erwin (Amsterdam: Elsevier) p 133
[5] M Muslihudin and D Hartini 2015 Perancangan Sistem Pendukung Pengambilan Keputusan Untuk Penerimaan Beasiswa Di SMA PGRI 1 Talang Padang Dengan Model Fuzzy Multiple Attribute Menggunakan Metode Simple Additive Weighting (SAW) Journal Technology Acceptance Model Vol 4 p 34 - 40.
[6] Depdiknas 2013 Peraturan Menteri Pendidikan Nasional No. 20, Tahun 2003 (Jakarta)
[7] A Nuraziq, Kusrini and E T Luthfi 2017 Rancang Bangun Sistem Pendukung Keputusan Penerimaan Beasiswa Menggunakan Metode Clustering K-Means Seminar Nasional Multi Disiplin Ilmu Vol 1
[8] J Aranda and W A G Natasya 2016 Penerapan Metode K-Means Cluster Analysis Pada Sistem Pendukung Keputusan Pemilihan Konsentrasi Untuk Mahasiswa International Class STMIK AMIKOM Yogyakarta Seminar Nasional Teknologi Informasi dan Multimedia vol IV no 2 (Yogyakarta: STMIK AMIKOM) p 1 - 6
[9] A Masruro, Kusrini and E T Luthfi 2015 Sistem Penunjang Keputusan Penentuan Lokasi Wisata Menggunakan K-Means Clustering dan TOPSIS Jurnal Ilmiah DASI Vol 15 No 4 p 1 – 5
[10] A Ristyawan, Kusrini and A Sunyoto 2015 Pemanfaatan Algoritma FCM Dalam Pengelompokan Kinerja Akademik Mahasiswa Konferensi Nasional Sistem dan Informatika 2015 (Bali)