



The Implementation of the K-Means Clustering Algorithm for Awarding Scholarships to Outstanding Students

Fahkri Adrian Syah¹, Syiah Putra Hasugian², Muhammad Vikhri Adafmi³, Abel Derosa Sibarani⁴, Lisnawita⁵

¹²³⁴⁵Informatics Engineering, Faculty of Computer Science, Universitas Lancang Kuning, Pekanbaru, Indonesia.

* Corresponding Author.

Email: lisnawita@unilak.ac.id

Received 01/06/2024, Revised 20/06/2024, Accepted 25/06/2024, Published 30/06/2024



This work is licensed under a [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/).

Abstract

Offering scholarships is an important tactic to promote higher education and scholastic success among students. The purpose of this study is to improve the impartiality and efficiency of scholarship awarding decision making by developing a decision support system built based on the K-Means clustering method. Students were grouped based on relevant variables and academic achievement requirements using the K-Means clustering algorithm. This approach creates homogeneous groupings based on scholarship recipients' history and academic performance data. The result of this clustering helps in recognizing trends and attributes that form the basis for future scholarship grant choices. This strategy was put into practice by creating a decision support system linked to student information and academic tracking. On processing the clustering data, cluster 0 with the status of eligible scholarship recipients amounted to 43 data and the cluster contained 157 data. It seems to contain data. Cluster 1 includes the status of scholarship recipients who do not meet the requirements. From the results of data analysis, it can be concluded that the scholarship recipient students are really outstanding students.

Keywords: Scholarship, Data mining, Clustering Algorithm

Introduction

Higher education in the current era is increasingly becoming a focal point of public attention as a primary foundation for producing quality human resources. One of the efforts to promote the quality of education is through the provision of scholarships to outstanding students. The selection process for scholarship recipients plays a crucial role in ensuring that scholarship funds are awarded to those who truly meet the criteria and need financial support (Fammaldo & Hakim, 2019).

However, the conventional scholarship selection process often faces challenges in managing large amounts of data and variables, both academic and non-academic. Therefore, the use of technology and data analysis methods becomes crucial to ensure objectivity and accuracy in determining scholarship recipients (Mardalius, 2018).



This research proposes the development of a decision support system using the K-Means Clustering algorithm to address these challenges (Sinawati & Syafiqoh, 2014). By analyzing student achievements based on certain criteria, such as Grade Point Average (GPA), selection test scores, and leadership experience, it is expected that this system can make a significant contribution to improving the efficiency and objectivity of the scholarship recipient selection process (Mardalius, 2018).

Through this approach, this research aims not only to enhance the efficiency of the selection process but also to provide a stronger basis for determining deserving scholarship recipients. Thus, this research is expected to make a real contribution to the development of more effective and data-oriented scholarship policies, creating a reliable and transparent decision support system (Sinawati & Syafiqoh, 2014).

K-Means is a method used to optimize datasets in classification (clustering). Clustering is the process of classifying data into several groups. Here, data from one group have similarities and differences in characteristics with data from other groups (Mardalius, 2018). The advantage of the K-means clustering method is that it is the most widely used and the easiest to use. The clustering results of the master data that meet the requirements of the BETUNAS scholarship recipients are the results of the grouping calculation of the recipient's suitability (Manihuruk et al., 2020). The aim of this research is to apply the K-means algorithm to minimize the influence of the grouping process suitability. This research uses several criteria: GPA, transcript scores, selection test scores, and achievements in both academic and non-academic fields.

One of the studies conducted using the K-Means algorithm by Sabrina Aulia Rahmah and Jovi Antares. The researchers used the Rapidminer algorithm to determine the impact of grants based on data from 80 applicants. Through attribute selection, K-means searches to capture all data into the current cluster. The results of the processing are accepted by 16%, considered by 61%, and rejected by 23% (Fathoni et al., 2021). The final analysis was conducted by Angolo Eko Wikaxono. This research develops the concept of data mining using the K-Means Clustering algorithm to predict which students will receive scholarships. This algorithm groups data into several clusters so that data with similar characteristics are grouped into similar clusters and data with different characteristics are grouped into different clusters (Wicaksono, 2017). Next, we study "Evaluation of Production Potential in Eastern Province (Case Study: Food Crop and Horticulture Services) Using K-Clustering" (Mahulae, 2020). Investigate "Clustering of Accessory Sales Data Using K-Means Algorithm Using Rapid Miner Studio 8.2" (Mardalius, 2018). We studied "Geographic Information System for Mapping the Poor in Kalambeski Regency Utilizing the K-means Clustering Method" (Hasymi et al., 2021). "Cendiawan Yayasan AMIK Tunas Bangsa: Application of SMART Method in Decision Making" (Andani, 2019). Research "Application of C4.5 Decision Tree Algorithm for University Teaching Staff Selection" (Hijriana & Rasyidan, 2017).

From this background, research will be conducted on the use of data mining to facilitate the selection process of prospective scholarship recipients. Any technique applied, such as data mining clustering using the K-Means algorithm, will be used in this study to address existing problems and allow the collection of master student data that can represent both academic and non-academic environments.

Materials and Methods

The steps taken to make recommendations regarding the data of prospective scholarship recipients require a procedure for sorting data and making predictions of scholarship member scores shown in Figure 1.

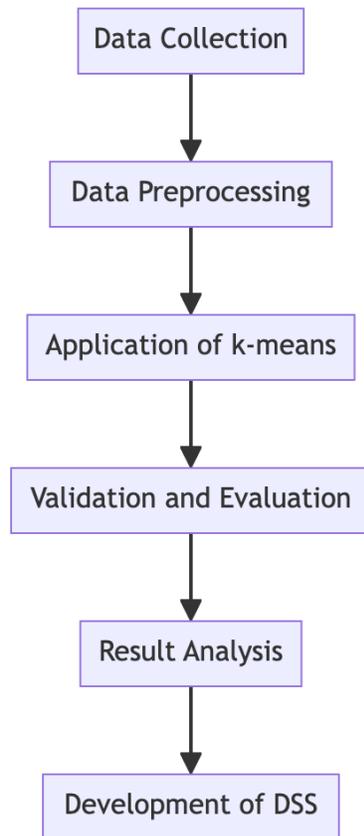


Fig 1. Phases of Research

Data Collection

The required data will be collected using various sources, including student academic records, leadership experience information data. The data will include GPA values, as well as the type and duration of leadership experience.

Data Preprocessing

The collected data will go through a preprocessing stage to clean up incomplete or invalid data. This preprocessing process includes normalizing the values, handling missing data, and converting the data to a format suitable for further analysis.

Application of K-Means Clustering Algorithm

K-Means Clustering algorithm will be applied to group students based on relevant variables. This grouping will form clusters that represent the level of similarity between students based on academic and non-academic achievements.



Validation of Results and Model Evaluation

The clustering results will be validated to ensure the quality and relevance of the resulting clusters. Furthermore, the performance of the system will be evaluated by comparing it with traditional scholarship selection methods through metrics such as prediction accuracy and execution time.

Analysis and Interpretation of Results

The results of the clustering will be interpreted and analyzed to gain a deeper insight into the patterns identified. This will help in understanding the characteristics of students in each cluster and the implications for scholarship selection policies.

Decision Support System Development

Based on the results of the analysis, a decision support system will be developed. This system will provide membership values for each student in certain clusters, providing an objective basis for determining scholarship recipients.

Through these methodological steps, it is hoped that this research can make a real contribution to the development of a decision support system for the selection of scholarship recipients that is more efficient and data-based.

Research Instruments

Research instruments are tools for collecting data or changing the subject matter of a particular research variable. For this reason, it is necessary to use a tool that can process relevant data accurately. Provide valid and consistent (reliable) research data (Siregar, 2019). This research uses several research tools, including:

1. Software One of the software techniques used in this research is the RapidMiner software technique used to combine datasets consisting of data. BETUNAS prospective undergraduate students.
2. This research uses a dataset of BETUNAS scholarship applicants who excel in 2022 as much as 200 data that undergoes a cleaning and transformation process consisting of several variables, such as GPA scores, certificates, selection exam results, and achievements achieved in academic and non-academic environments. This process results in data grouping based on the similar characteristics of each data, allowing you to find patterns and hidden information in your data.

Results and Discussion

The implementation of this research uses the calculation of the K-Means algorithm which is carried out using the Rapid Miner Studio tool. Rapid Miner is a tool that allows users to perform calculations easily using operators. The operator is in charge of modifying the data and is connected to the operator node (Astuti, 2017). While the user's job is to connect these nodes and display the results of the calculation graph.

From the research process carried out, the results obtained from 200 datasets of student lists of BETUNAS scholarship applicants in 2022. there are 43 data following cluster 0 which means that students qualify as scholarship recipients or receive scholarships while 157 other data follow cluster 1. This means that they do not meet cluster 1 requirements as scholarship recipients.

Table 1. BETUNAS registrant dataset

Number	Name	Grade point Average	Numeric Grade
1	Devi Miswiyani Agusti Ningsih	4,00	80
2	Ida Holidia	4,00	80
3	Santi Gustin	4,00	80
...
299	Annisaa Ambarnis	3,84	80
200	Juliani Putri Ahmad	3,84	80

Data Processing with Rapid Miner

1. Data clustering in the study was carried out using the following software, Rapidminer:
2. Open the rapid miner studio software version.
3. Then select the File menu and select New Process.
4. Click on the panel, select "read Excel", drag the operator, and it will appear in the process panel and can be used to import Excel data. This includes data transformation and data cleaning processes.
5. The next step is to enter the model that will be used for the data mining clustering process by entering the operator menu then selecting K-Means.
6. Next, in the clustering process, first determine the number of clusters to $K = 2$ and the mixed measure to be mixed Euclidean distance.
7. The next step is to connect the connectors of each process to the main process and click the Run button to start clustering.

Clustering results using K-means algorithm

a. Data View

After passing the stages of the K-Means Clustering process using the Rapidminer Studio application, the next process is to display the results of the application. In Figure 2, the data view of the clusterization results that have been obtained will be displayed as follows:

Loop Parameters (9 rows, 3 columns)

iteration	k-mean...	Davies ...
1	2	0.006
7	8	0.228
5	6	0.231
2	3	0.177
3	4	0.249
8	9	0.219
4	5	0.241
6	7	0.233
9	10	0.214

Fig 2. View Cluster Results

b. Cluster Model

The results are presented in the form of test results using Rapidminer software on a data set of 200 prospective scholarship recipient data, thus forming two clusters. There are 43 data in Cluster 0 (first cluster) and 157 data in Cluster 1 (second cluster).

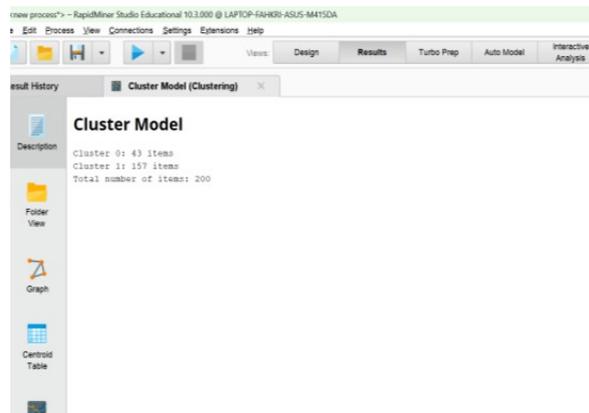


Fig 3. Cluster Model Results

c. Display Folder

Folders containing students who are eligible to receive the Indonesia Pintar Scholarship Program can be seen in Figure 4 below. Cluster 0 (first cluster) has 43 student names that are included in the scholarship recipient category.

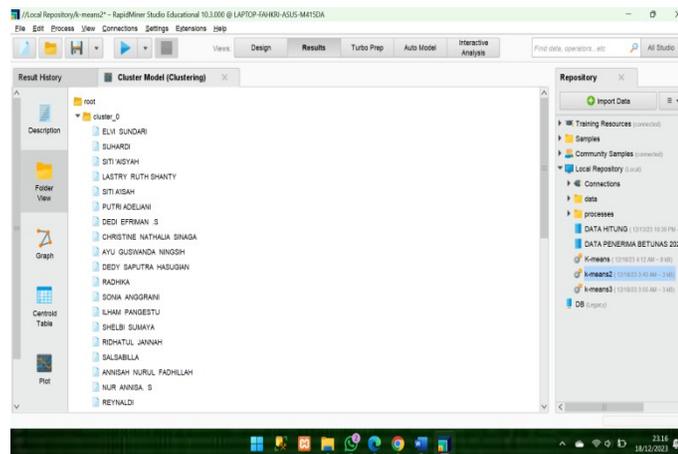


Fig 4. Result of Folder View

Conclusion

Based on the research findings, it can be concluded that the K-means clustering algorithm applied to student data shows the efficacy of the program for underprivileged children. The data mining process produces a set of scholarship recipients based on three research variables, making it easier for related parties to identify BETUNAS recipients using data mining techniques. Based on the results of data mining, there are two clusters. at the time of grouping data processing, cluster 0 with the status of eligible scholarship recipients contains 43 data and cluster 1 contains 157 data cluster 1 includes the status of scholarship recipients who do not meet the requirements from the results of data analysis it can be concluded that the scholarship recipient students are really outstanding students.

References

- E. Fammaldo and L. Hakim, "Penerapan Algoritma K-Means Clustering Untuk Pengelompokan Tingkat Kesejahteraan Keluarga Untuk Program Kartu Indonesia Pintar," *J. Ilm. Teknol. Infomasi Terap.*, vol. 5, no. 1, pp. 23–31, 2019, doi: 10.33197/jitter.vol5.iss1.2018.249.
- B. G. Sudarsono and S. P. Lestari, "Clustering Penerima Beasiswa Yayasan Untuk Mahasiswa Menggunakan Metode K-Means," *J. Media Inform. Budidarma*, vol. 5, no. 1, p. 258, 2021, doi: 10.30865/mib.v5i1.2670.
- U. Syafiqoh and S. Informasi, "Sistem Pendukung Keputusan Penentuan Pemberian Bantuan Biaya Pendidikan Menggunakan Algoritma K-Means," pp. 37–42, 2014.
- N. A. Manihuruk, M. Zarlis, E. Irawan, and H. S. Tambunan, "Penerapan Data Mining Dalam Mengelompokkan Calon Penerima Beasiswa Dengan Menggunakan Algoritma K-Means," *KOMIK (Konferensi Nas. Teknol. Inf. dan Komputer)*, vol. 4, no. 1, pp. 29–34, 2020, doi: 10.30865/komik.v4i1.2575.
- Fathoni, M. Y., Darmansah, D., & Januarita, D. (2021). Sistem Pendukung Keputusan Pemilihan Siswa Teladan Menggunakan Metode Simple Additive Weighting (SAW) Pada SMK Telkom Purwokerto. *Jurnal Sistem Informasi Dan Komputer/Jurnal Sisfokom*, 10(3), 346–353. <https://doi.org/10.32736/sisfokom.v10i3.1202>.
- Wicaksono, A. E. (2017). Implementasi Data Mining Dalam Pengelompokan Data Peserta Didik Di Sekolah Untuk Memprediksi Calon Penerima Beasiswa Dengan Menggunakan Algoritma K-Means (Studi Kasus Sman 16 Bekasi). 21(3). <https://ejournal.gunadarma.ac.id/index.php/tekno/article/view/1599>



- H. Mahulae, “Pengelompokan Potensi Produksi Buah-Buahan di Provinsi Sumatera Utara dengan Menerapkan K-Clustering (Studi Kasus : Dinas Tanaman Pangan dan Holtikultura),” *JURIKOM (Jurnal Ris. Komputer)*, vol. 7, no. 2, p. 312, 2020, doi: 10.30865/jurikom.v7i2.2122.
- M. Mardalius, “Pemanfaatan Rapid Miner Studio 8.2 Untuk Pengelompokan Data Penjualan Aksesoris Menggunakan Algoritma K-Means,” *Jurteksi*, vol. 4, no. 2, pp. 123–132, 2018, doi: 10.33330/jurteksi.v4i2.36.
- M. Ali Hasymi, A. Faisol, and F. Ariwibisono, “Sistem Informasi Geografis Pemetaan Warga Kurang Mampu Di Kelurahan Karang Besuki Menggunakan Metode K-Means Clustering,” *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 5, no. 1, pp. 284–290, 2021, doi: 10.36040/jati.v5i1.3269.
- S. R. Andani, “Penerapan Metode SMART dalam Pengambilan Keputusan Penerima Beasiswa Yayasan AMIK Tunas Bangsa,” *J. Sist. dan Teknol. Inf.*, vol. 7, no. 3, p. 166, 2019, doi: 10.26418/justin.v7i3.30112.
- N. Hijriana and M. Rasyidan, “Penerapan Metode Decision Tree Algoritma C4.5 Untuk Seleksi Calon Penerima Beasiswa Tingkat Universitas,” *Sains Dan Teknol.*, vol. 3, no. 1, pp. 9–13, 2017.
- Siregar, A. M. (2019). Penerapan Algoritma K-Means Untuk Pengelompokan Daerah Rawan Bencana Di Indonesia. *Internal*, 1(2), 1–10. <https://doi.org/10.32627/internal.v1i2.42>.
- Astuti, F. D. (2017). Penerapan Data Mining Untuk Clustering Data Penduduk Miskin Menggunakan Algoritma Hard C-Means. *Data Manajemen Dan Teknologi Informasi (DASI)*, 18(1), 64–69. <https://eprints.akakom.ac.id/6058/>