Optimizing the Selection of KIP Scholarship Recipients at STKIP Al Maksum Langkat using K-Means Clustering Method

by Universitas Potensi Utama

Optimizing the Selection of KIP Scholarship Recipients at STKIP Al Maksum Langkat using K-Means Clustering Method

Muhammad Hari Hasibuan[1*], Wanayumini[1], Rika Rosnelly[1]

[1]Master of Computer Science Study Program, Computer Science, University of Potensi Utama Medan
JL. KL. Yos Sudarso Km. 6,5 No. 3-A, Kota Medan, Sumatera Utara, Indonesia
*Email: mharyhsb28@gmail.com, wanayumini@gmail.com, rikarosnelly@gmail.com

ABSTRACT

The KIP scholarship program is a form of government assistance to support the education of underprivileged students in Indonesia. STKIP Al Maksum Langkat is one of the institutions that provides the KIP scholarship program to its students. However, the selection process for scholarship recipients is still done manually and can be time-consuming and less effective. In this study, the K-means Clustering method was applied to optimize the selection of KIP scholarship recipients at STKIP Al Maksum Langkat. This method was used to group scholarship recipient data based on several variables, such as family income, number of dependents, and socio-economic background. The results showed that the K-means Clustering method can effectively group scholarship recipients based on their characteristics and assist in selecting eligible scholarship recipients. Additionally, this method can improve the effectiveness and efficiency of the scholarship selection process at the institution. Overall, this study demonstrates that the K-means Clustering method can be an effective approach to optimize the selection of KIP scholarship recipients at STKIP Al Maksum Langkat.

Keywords : Optimizing, KIP Scholaship Recipients, K-Mean Clustering Method

Medan, Sumatera Utara, Indonesia 26th January 2023
Faculty of Economics and Business Universitas Muhammadiyah Sumatera Utara

2697

INTRODUCTION

Higher education in Indonesia should be able to shape individuals with innovative, creative, independent, and highly competitive characters through the mastery of the latest knowledge and technology. Through the Indonesia Pintar Program, the Indonesian government has issued the Indonesia Pintar Kuliah Card (KIP Kuliah) since 2020 as one form of educational assistance provided to high school graduates from underprivileged families to continue their education to higher education or academy level.

Jaroji's research on selecting Bidikmisi scholarship recipients at Polbeng found that the K-Means algorithm grouped candidates into four clusters. Cluster 0 recommended with consideration, while clusters 1, 2, and 3 provided highly eligible, eligible, and less eligible recommendations, respectively. The clustering results serve as recommendations for decision makers and not as the ultimate determinant of scholarship recipients.

Sudarsono conducted a research on determining the group of scholarship recipients using the K-Means method. This method is used for data clustering with K-Means algorithm and Euclidean model. The result shows one group based on the proximity distance to the data value, however, this only provides recommendations and does not definitively determine scholarship recipients. The K-Means algorithm was introduced by J.B. MacQueen in 1976. This method partitions data into clusters so that data with similar characteristics are grouped into the same cluster, while data with different characteristics are grouped into different clusters.

The use of the k-means method for other analytical purposes has been demonstrated in a study published by (Ghofar & Kurniawan, 2018). In this paper, the authors present the results of an analysis of customer data from UMS Store using a sample size of 357. The k-means method was used to cluster customer data into 3 clusters. After comparing the attribute values in each cluster, it was concluded that potential customers were found in cluster 3. Therefore, cluster 3 was recommended as the target market for UMS Store.

STKIP Al Maksum Langkat is a private college in Langkat Regency that offers scholarships for students from lower-middle economic backgrounds. However, the scholarship selection process at STKIP Al Maksum Langkat faces challenges in deciding which students are eligible for scholarships. This is because the evaluation process is not always based on accurate calculations but is more influenced by the decision-makers who ultimately determine the scholarship recipients. Therefore, an application with a more objective method is needed to predict potential scholarship recipients.

PROPOSED FEATURE: OPTIMIZING K-MEANS CLUSTERING METHOD

K-Means clustering is a cluster analysis method that aims to divide objects into k clusters, where each cluster is obtained through the nearest mean observation. Algoritma K-Means is an evolutionary algorithm whose operation method has a similar meaning to its name. The K-Means algorithm is a clustering technique used in Data Mining modeling process that is performed without supervision and groups data into clusters. Each cluster has similar characteristics to its members, and the goal is to minimize differences between data points within a cluster while maximizing differences with other clusters. The algorithm utilizes several terms, including:

1. Cluster, which refers to a group or collection.
2. Centroid, which represents the central point used to determine the distance between data points.
3. Iteration, which means repeating the process until the result has converged.

Medan, Sumatera Utara, Indonesia 26th January 2023
Faculty of Economics and Business Universitas Muhammadiyah Sumatera Utara

2698

E-ISSN 2985-7910

Proceeding Medan International Conference Economics and Business

Volume 1, Year 2023
"Entrepreneurship on Global Economics Development in the Era of Society 5.

The general steps of the K-Means algorithm include:
1. Determining the number of clusters, k.
2. Randomly selecting k objects to be used as centroid points.
3. Calculating k centroids or central points.
4. Grouping objects to their nearest centroid based on the Euclidean distance.

$$dik = \sqrt{\sum (Cij - Ckj)^2} \, n \; i=1$$

where:

Cij = centroid of cluster i

Ckj = data point j

5. Recalculate all centroid points.
6. Repeat steps 3-5 until the centroid values no longer change.

Clustering is a process of grouping a set of data with unknown class attributes, with the aim of minimizing similarities among the classes. For example, a set of objects can be clustered into several classes based on a certain cluster group. Cluster can be interpreted as a group, so clustering analysis will produce a number of clusters. In clustering analysis, a set of data already has similarities among its members, so each member with similar characteristics is grouped into one or more clusters. The goal of clustering data is to minimize the objective function by minimizing the variation within a cluster and maximizing the variation between clusters.

The Ministry of Education and Culture (Kemendikbud) has launched the Merdeka Belajar Episode IX Kartu Indonesia Pintar (KIP) Kuliah Merdeka program. The KIP Kuliah scholarship program aims to improve access to quality higher education for underprivileged communities, especially those with a lower middle-class status. With this program, it is hoped that more people in need will be able to obtain higher education that is more equitable and of better quality. This is in line with the vision and mission of President Joko Widodo to create a highly skilled workforce in Indonesia. It is hoped that in the future, Indonesia will have a qualified workforce that is ready to compete on the global stage.

DATA AND EXPERIMENTAL SETUP

K-Means Clustering is used to group data, and the clustering results are then refined using the Backpropagation Neural Network algorithm to determine the accuracy level of the previous data clustering. This is done as a consideration for determining KIP Scholarship recipients STKIP Al Maksum Langkat. The following are some stages and processes of clustering using the K-Means Clustering method:

1. Determining the Number of Clusters

In the K-Means Clustering algorithm, the first step is to determine the number of clusters to be used. In the system for determining recipients of the KIP Kuliah STKIP Al Maksum Langkat, 2 clusters will be used:
1. First Cluster (C0) = Accepted
2. Second Cluster (C1) = Rejected

At this stage, random centroid values will be determined for student data based on 13 criteria to determine KIP Scholarship recipients at STKIP Al Maksum Langkat. The 13 criteria are:
1. Father's Occupation (FO)

Medan, Sumatera Utara, Indonesia 26th January 2023
Faculty of Economics and Business Universitas Muhammadiyah Sumatera Utara

2699

MICEB 2023

E-ISSN 2985-7910

Proceeding Medan International Conference Economics and Business
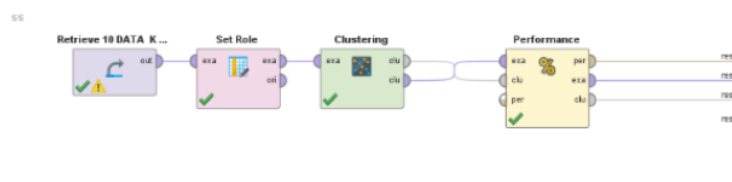
Volume 1, Year 2023
"Entrepreneurship on Global Economics Development in the Era of Society 5.

2. Father's Income (FI)
3. Father's Status (FS)
4. Mother's Occupation (MO)
5. Mother's Income (MI)
6. Mother's Status (MS)
7. Number of Dependents (ND)
8. Family Home Ownership (FHO)
9. Electricity Source Used by Family (ES)
10. Family Home Land Area (FLA)
11. Family Home Building Area (FBA)
12. Main Water Source (MWS)
13. Ownership of Bathroom and Toilet Facilities (BT)

Table 1. Initial Test Data

Centroid	Student Name	FO	FI	FS	MO	MI	MS	ND	FHO	ES	FLA	FBA	MWS	BT
1	Anggita Dwi Sukma	7	4	1	7	1	1	2	1	1	1	3	3	1
2	Marisa Umi Fahira	8	1	2	7	5	1	2	1	1	1	5	3	1
3	Mutia Anisa	7	4	1	7	1	1	2	1	1	1	2	3	1
4	Adinda Pratiwi	5	3	1	8	1	1	2	1	1	2	1	3	1
5	Ahmad Khairul Fakhri	7	10	1	8	1	1	2	1	1	4	5	3	1
6	Nadia Yasmina Fajra Tanjung	7	9	1	8	1	1	3	4	1	2	2	3	1
7	YUNITA	5	7	1	7	1	1	3	1	1	4	5	3	1
8	FIKRI	5	6	1	8	1	1	3	1	1	4	4	1	1
9	ZULFADLI	3	5	1	8	1	1	1	1	1	5	5	3	1
10	MUHAMMAD ANDI	7	5	1	8	1	1	4	1	1	5	5	3	2

The application of K-Means Clustering method is performed on 10 student data who are potential recipients of KIP Scholarship at STKIP Al Maksum Langkat using RapidMiner Studio. In addition, K-Means Clustering method visualization is also conducted using RapidMiner Studio to facilitate understanding of the analysis results. Figure 1 illustrates the operators used in the implementation of K-Means Clustering method with RapidMiner Studio as follows.


Figure 1. RapidMiner Process Display with K-Means Method

Medan, Sumatera Utara, Indonesia 26th January 2023
Faculty of Economics and Business Universitas Muhammadiyah Sumatera Utara

2700

![MICEB 2023]

E-ISSN 2985-7910

Proceeding Medan International Conference Economics and Business

Volume 1, Year 2023
"Entrepreneurship on Global Economics Development in the Era of Society 5.

The following is the result of the analysis using RapidMiner Studio application with K-Means Clustering method. The display can be seen in the following figure 2:

Ro...	id	Nama Siswa	cluster	PA	PHA	SA	PI	PHI	SI	JT	KR	SL	LT	LB	SAU	MCK
1	1	Anggita Dwi ...	cluster_0	7	4	1	7	1	1	2	1	1	1	3	3	1
2	2	Marisa Umi F...	cluster_0	8	1	2	7	5	1	2	1	1	1	5	3	1
3	3	Mutia Anisa	cluster_0	7	4	1	7	1	1	2	1	1	1	2	3	1
4	4	Adinda Pratiwi	cluster_0	5	3	1	8	1	1	2	1	1	2	1	3	1
5	5	Ahmad Khair...	cluster_1	7	10	1	8	1	1	2	1	1	4	5	3	1
6	6	Nadia Yasmi...	cluster_1	7	9	1	8	1	1	3	4	1	2	2	3	1
7	7	YUNITA	cluster_1	5	7	1	7	1	1	3	1	1	4	5	3	1
8	8	FIKRI	cluster_1	5	6	1	8	1	1	3	1	1	4	4	1	1
9	9	ZULFADLI	cluster_1	3	5	1	8	1	1	1	1	1	5	5	3	1
10	10	MUHAMMAD ...	cluster_1	7	5	1	8	1	1	4	1	1	5	5	3	2

Figure 2. Result of K-Means Clustering Process using RapidMiner Studio

In figure 2, the result of the K-Means Clustering process using RapidMiner is shown. The result is the same as the manual K-Means Clustering process. There are 4 students in Cluster 0 (accepted) and 6 students in Cluster 1 (rejected).

The following are the results of the application of K-Means Clustering method based on 10 student data registered in the KIP Kuliah portal. There were 4 student data included in category C0 (Accepted) and 6 student data included in category C1 (Not Accepted). The visualization of the Cluster Model can be seen in Figure 3 below

Cluster Model

Cluster 0: 4 items
Cluster 1: 6 items
Total number of items: 10

Figure 3. Visualization of Cluster Model from 10 Student Data

Based on the calculation results of the 10 student data, there are 40% of students who belong to Cluster 0 (accepted) category and 60% of students belong to Cluster 1 (not accepted) category.

PerformanceVector

```
PerformanceVector:
Avg. within centroid distance: -10.025
Avg. within centroid distance_cluster_0: -8.438
Avg. within centroid distance_cluster_1: -11.083
Davies Bouldin: -1.068
```

Figure 4. PerformanceVector

Medan, Sumatera Utara, Indonesia 26th January 2023
Faculty of Economics and Business Universitas Muhammadiyah Sumatera Utara

2701

MICEB 2023

E-ISSN 2985-7910

Proceeding Medan International Conference Economics and Business

Volume 1, Year 2023
"Entrepreneurship on Global Economics Development in the Era of Society 5.

PerformanceVector is a set of metrics used to evaluate the quality of clustering results. The PerformanceVector includes several measures, such as the average within centroid distance, which is the average distance of each point in a cluster to its centroid. The PerformanceVector also includes the average within centroid distance for each individual cluster, which provides insight into the variability within each cluster.

In addition, the PerformanceVector includes the Davies-Bouldin Index, which measures the similarity between clusters based on the ratio of the distance between their centroids and the average distance within each cluster.

In this specific case, the PerformanceVector shows an average within centroid distance of -10.025, with a lower distance indicating a better clustering result. The PerformanceVector also shows that the average within centroid distance for Cluster 0 is -8.438 and for Cluster 1 is -11.083, indicating that Cluster 0 has a tighter grouping of points around its centroid than Cluster 1. The Davies-Bouldin Index is -1.068, which is considered to be a good clustering result as lower values indicate better separation between clusters.

RESULT AND DISCUSSION

The variables used in this study are sourced from the registration data of KIP Kuliah students in 2020, which comes from the Kemenristek DIKTI database channel Sim KIP Kuliah. The response variable (Y) in this study is the status of receiving KIP Kuliah scholarship at STKIP Al Maksum Langkat. Meanwhile, the predictor variable (X) uses micro data that contains characteristics of prospective KIP Kuliah scholarship applicants at STKIP Al Maksum Langkat. Details of the response variable and predictor variable of micro data can be seen in Table 2.

Table 2. The variables used in this study are sourced
from the registration data of KIP Kuliah students in 2020.

Variable	Variable Name	Description	
Y	Status of KIP Scholarshi Acceptance	0 = Accepted 1 = Rejected	
X1	Father's Occupation	1 = Civil Servant 2 = Private Employee 3 = Entrepreneur 4 = Military/Police	5 = Farmer 6 = Fisherman 7 = Other 8 = Unemploye
X2	Father's Income	1 = No Income 2 = < Rp.250.000 3 = Rp. 250.000 4 = Rp. 500.001 5 = Rp. 750.000	6 = Rp. 1.000.000 7 = Rp. 1.500.000 8 = Rp. 1.750.000 9 = Rp. 2.000.000
X3	Father's Status	1 = Still alive 2 = Deceased 3 = Divorced	
X4	Mother's Occupation	1 = Civil Servant 2 = Private Employee 3 = Entrepreneur 4 = Military/Police	5 = Farmer 6 = Fisherman 7 = Other 8 = Unemploye
X5	Mother's Income	1 = No Income 2 = < Rp.250.000 3 = Rp. 250.000 4 = Rp. 500.001	6 = Rp. 1.000.000 7 = Rp. 1.500.000 8 = Rp. 1.750.000 9 = Rp. 2.000.000
X6	Mother's Status	1 = Still alive 2 = Deceased	

Medan, Sumatera Utara, Indonesia 26th January 2023
Faculty of Economics and Business Universitas Muhammadiyah Sumatera Utara

2702

MICEB 2023

E-ISSN 2985-7910

Proceeding Medan International Conference Economics and Business

Volume 1, Year 2023

"Entrepreneurship on Global Economics Development in the Era of Society 5.

		3 = Divorced	
		1 = 1 person	4 = 4 people
X7	Number of Dependents	2 = 2 people	5 = More than 5 people
		3 = 3 people	6 = No Dependents
		1 = Owned	4 = Squatting
X8	Family's Residential	2 = Annual Rent	5 = Not Owned
		3 = Monthly Rent	6 = Unauthorized Occupancy
		1 = PLN	
X9	Main Electricity Source Used by the Family	2 = Generator/Self-generated	
		3 = Solar Energy	
		4 = Relying on Neighbor	
		1 = > 200 M²	4 = 25-50 M²
X10	Size of Family's Residential Land	2 = 100 – 200 M²	5 = < 25 M²
		3 = 50 - 99 M²	
		1 = > 200 M²	4 = 25-50 M²
X11	Size of Family's Residential Building	2 = 100 – 200 M²	5 = < 25 M²
		3 = 50 - 99 M²	
		1 = Bottled Water	
X12	Main Water Source	2 = PDAM (Public Water Utility)	
		3 = Well	
X13	Ownership of Bathroom Facilities	1 = Owned, Inside the house	
		2 = Owned, Outside the house	
		3 = Shared with others	

Table 3. The Final Results

No	Student Name	Cluster
1	Anggita Dwi Sukma	Cluster 0 (Accepted)
2	Marisa Umi Fahira	Cluster 0 (Accepted)
3	Mutia Anisa	Cluster 0 (Accepted)
4	Adinda Pratiwi	Cluster 0 (Accepted)
5	Ahmad Khairul Fakhri	Cluster 1 (Rejected)
6	Nadia Yasmina Fajra Tanjung	Cluster 1 (Rejected)
7	YUNITA	Cluster 1 (Rejected)
8	FIKRI	Cluster 1 (Rejected)
9	ZULFADLI	Cluster 1 (Rejected)
10	MUHAMMAD ANDI	Cluster 1 (Rejected)

Table 3 shows the final result of the clustering of the KIP scholarship recipients at STKIP Al Maksum Langkat. It can be seen that there are 4 students who are eligible to receive the KIP scholarship, while 6 students are not eligible.

Medan, Sumatera Utara, Indonesia 26th January 2023
Faculty of Economics and Business Universitas Muhammadiyah Sumatera Utara

2703

CONCLUSION

The journal "Optimizing the Selection of KIP Scholarship Recipients at STKIP Al Maksum Langkat using K-Means Clustering Method" utilizes the K-Means clustering method to optimize the selection of KIP scholarship recipients at STKIP Al Maksum Langkat. Based on the 10 data used, the results indicate that 40% of participants are eligible to receive the KIP scholarship, while 60% do not meet the requirements to receive the scholarship. However, it is important to note that this conclusion is based on limited data, and therefore the results may not be representative of a larger or different population.

REFERENCES

Kemendikbud, "Pedoman Pendaftaran Kartu Indonesia Pintar Kuliah (KIP-K)," Puslapdik, vol. 1, no. 1, pp. 1–18, 2021, [Online]. Available: https://kip-kuliah.kemdikbud.go.id/uploads/BsImnu09yFOxop5dfJAwkaRleMTUqP_tgl20200412205459.pdf

H. Pamungkas, "Implementasi Metode K-Means Dalam Penentu Keputusan Penerimaan Beasiswa Di Smk Negeri 1 Lengkong," vol. 01, no. 11, p. 10, 2017.

H. S. Tambunan, I. Gunawan, and S. Sumarno, "Prediksi Jumlah Pendapat Beasiswa PPA dan BBP Menggunakan Jaringan Syaraf Tiruan Backpropagation," J. Media Inform. Budidarma, vol. 3, no. 4, p. 346, 2019, doi: 10.30865/mib.v3i4.1327.

N. Fitria Hastuti, "Pemanfaatan Metode K-Means Clustering dalam Penentuan Penerima Beasiswa," Skripsi, 2016.

M. A. Ghofar and Y. I. Kurniawan, "APLIKASI PENGELOMPOKAN PELANGGAN PADA UMS STORE MENGGUNAKAN ALGORITMA K-MEANS , Program Studi Informatika , Fakultas Komunikasi dan Informatika," vol. 4, no. 1, 2018.

M. Sadikin, T. A. Hapsari, F. I. Komputer, U. M. Jakarta, D. Khusus, and I. Jakarta, "Analisis Pola Keterkaitan Profil Dengan Tingkat Kehadiran Pegawai Menggunakan Clustering K-Means . Studi Kasus Pada Sekretariat BPJT The Analysis of Association Rule Between Profile and Employee ' s Presence Using K-Means Clustering . Case Study BPJT Secr," pp. 104–112, 2019.

I. Kamila, U. Khairunnisa, and M. Mustakim, "Perbandingan Algoritma K-Means dan K-Medoids untuk Pengelompokan Data Transaksi Bongkar Muat di Provinsi Riau," J. Ilm. Rekayasa dan Manaj. Sist. Inf., vol. 5, no. 1, p. 119, 2019, doi:

K. I. P. Kuliah, D. Metode, and M. S. Sompa, "Clustering Tingkat Ekonomi Mahasiswa Calon Penerima Kartu Indonesia Pintar Program Sarjana," vol. 1, no. 2, pp. 65–71, 2022.

"Tampilan PENERAPAN MACHINE LEARNING UNTUK PENENTUAN SEGMENTASI MAHASISWA BARU DENGAN METODE K MODES.pdf."

R. K. Dinata, N. Hasdyna, S. Retno, and M. Nurfahmi, "K-means algorithm for clustering system of plant seeds specialization areas in east Aceh," Ilk. J. Ilm., vol. 13, no. 3, pp. 235–243, 2021, doi: 10.33096/ilkom.v13i3.863.235-243.

Medan, Sumatera Utara, Indonesia 26th January 2023
Faculty of Economics and Business Universitas Muhammadiyah Sumatera Utara

2704

Optimizing the Selection of KIP Scholarship Recipients at STKIP Al Maksum Langkat using K-Means Clustering Method