Determining Study Groups Based on Student Profile Criteria Using K-Means Method

Denny Kurniadi[1]*, Rahmadona Safitri[2], Dedy Irfan[3], Khairi Budayawan[4]
[1234]Pendidikan Teknik Informatika, Universitas Negeri Padang, Indonesia
Corresponding Author: dennykurniadi@ft.unp.ac.id

INTISARI

Mahasiswa Prodi PTI FT-UNP disusun ke dalam rombongan belajar berdasarkan satu kriteria yaitu jalur masuk dan dilakukan pada mahasiswa baru. Akibatnya, suatu rombongan belajar yang terbentuk memiliki perbedaan yang jauh antara satu rombongan dengan rombongan lainnya menyangkut kemampuan akademik, asal daerah, asal sekolah dan factor lainnya. Hal ini tidak menguntungkan bagi minat dan motivasi belajar mahasiswa dan pengajar juga mengalami kesulitan dalam menentukan strategi belajar antar rombongan belajar mahasiswa. Perbaikan penyusunan rombongan belajar dapat dilakukan dengan penetapan profil mahasiswa sebagai dasar penyusunan. Melalui penerapan metode K-Means yang diimplementasikan dalam suatu aplikasi, pengklusteran mahasiswa yang dijadikan dasar distribusi pembentukan rombongan belajar terbukti menjadi lebih baik. Rombongan belajar yang dihasilkan berhasil memperlihatkan rombongan belajar yang bersifat heterogen antar anggotanya dan homogen disetiap rombongan belajar.

Kata kunci: k-means clustering, grouping criteria, student profile.

ABSTRACT

Students of PTI FT-UNP Study Program are arranged into study groups based on one criterion, namely the entry route and is carried out on new students. As a result, a study group that is formed has far differences between one group and another regarding academic ability, regional origin, school origin and other factors. This is not beneficial for the interest and motivation to learn students and teachers also have difficulty in determining learning strategies between student study groups. Improvements in the preparation of study groups can be done by determining student profiles as the basis for preparation. Through the application of the K-Means method which is implemented in an application, the student clustering which is used as the basis for the distribution of the formation of study groups is proven to be better. The resulting study group succeeded in showing a heterogeneous study group among its members and homogeneous in each study group

Keywords: k-means clustering, grouping criteria, student profile.

INTRODUCTION

Students of PTI FT-UNP Study Program are grouped into study classes based on the entrance route. That is, the student study group is a form of homogeneous grouping[3][4].

Homogeneous and systematic grouping of students does not provide favorable learning conditions for [8]. Based on the results of interviews with several lecturers who teach at the Department of Electronic Engineering, they said that lecturers also had difficulties in determining the techniques or methods of classroom learning.

Based on the abundance of student data and data on the number of student graduations, hidden information can be found by processing student data so that it is useful for the university [1].

Student data processing needs to be done to find out important information in the form of new knowledge (knowledge discovery), for example information on classifying student data based on

profiles and academic data. This new knowledge can help universities to classify students' graduation rates in order to determine strategies to increase graduation in the following years.

In classifying student learning classes, student profiles are determined which are used as criteria in the preparation of learning classes. The profile is in the form of entry point, gender, GPA, school origin, regional origin, GPA, and the level of parents' economic ability [6].

Heterogeneous class is a class in which there is a mixture of student characteristics [3]. In heterogeneous classes, educators provide opportunities for students [3] including (a) meeting with class members with various characteristics, (b) the emergence of an attitude of helping, tolerance, mutual respect, empathy, willing to recognize the strengths of others and (c) students The slow learners are motivated to keep up with the fast learners.

There are several things that affect learning motivation, namely (a) the economic conditions of parents [6] and (b) the adaptability among students [5].

To create a good learning atmosphere and increase student learning motivation, student study groups were arranged in the form of heterogeneous classes [5].

In order for the management of the Department of Electronics Engineering to have no difficulty, an application for the preparation of learning classes was made by applying data mining using a clustering algorithm. [1] The criteria that will be used are entry point, GPA, gender, school origin, regional origin and parents' economic level. This application can also be used for the preparation of student study groups at each turn of the semester.

METHOD

K-Means Clustering
K-Means is one of the algorithms in data mining that is commonly used to cluster data. There are many approaches to creating clusters, including making rules that dictate membership in the same group based on the level of equality among its members (Winardi, 2016) [1]. The closer the distance, the higher the similarity. The calculation steps with the k-means clustering algorithm:

1. Determine the number of clusters, where the number of clusters is K=3.
2. Determine the centroid value.
3. Calculating the distance between the data and the center of the centroid using the Ecludiean distance formula which is formulated as follows:

$$D\,(i,j) = \sqrt{(X_{1i} - X_{1j})^2 + (X_{2i} - X_{2j})^2 + ... + (X_{ki} - X_{kj})^2}\quad(1)$$

$D\,(i,j)$ = Distance of data to i to cluster center j
X_{ki} = Data to i on attribute data to k
X_{kj} = Center point to j on attribute k
4. Allocate each data to the nearest centroid.

K-Means Clustering Design
Before performing the clustering calculations, it is necessary to group the criteria and the value of each of these criteria.

Determination of Criteria
The following criteria will be used to cluster student data:

Table 1. Description of Student Criteria

Criteria	Criteria Type		Value
Jalur Masuk	1.	SNMPTN	SNMPTN = 1
	2.	SBMPTN	SBMPTN = 2
	3.	Mandiri	Mandiri = 3
Asal Sekolah	1.	SMA/MA	SMA/MA = 1
	2.	SMK	SMK = 2
	3.	MA	
Jenis Kelamin	1. Male		M= 1
	2. Female		F = 2
IPK	IPK		IPK
Penghasilan Orang Tua	1. < Rp. 750.000		1
	2. Rp. 750.001 – Rp. 1000.000		2
	3. Rp. 1.000.000 – Rp. 2.500.000		3
	4. Rp. 2.500.001 – Rp. 4.500.000		4
	5. Rp. 4.500.001 – Rp. 5.500.000		5
	6. > Rp. 5.500.000		6
Asal Daerah	In West Sumatra		1
	Out West Sumatra		2

Determination of the Number of Clusters
The k-means clustering algorithm was used to calculate the level of heterogeneity of PTI students. So, there are three clusters, namely cluster_0 for high-level homogeneity, cluster_1 for moderate-level homogeneity and cluster_2 for low-level homogeneity.

Class Distribution Flowchart
This flowchart describes the algorithm for distributing clustering results to form eight learning class groups.
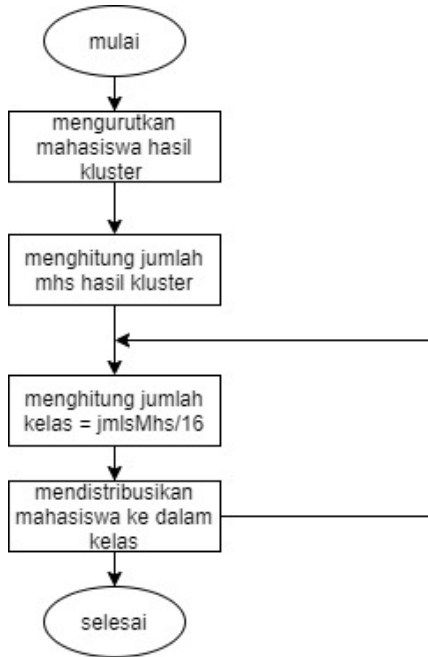
Figure 1. Class Distribution Flowchart

Determination of Centroid Value

The centroid value is usually determined randomly. The following centroid values are used:

Table 2. Centroid

Criteria	C1	C2	C3
IPK	3.31	3.32	3.31
k_jk	1.06	1.81	1.53
k_jm	2.51	2.68	1.0
k_as	1.65	2.32	1.38
k_ad	1.367	1.35	1.38
Atribut	3.31	3.32	3.31

RESULTS AND DISCUSSION

Login Page
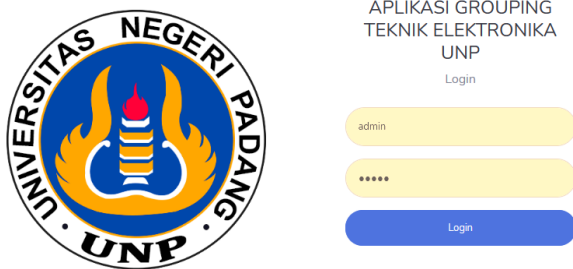
Figure 2. Login Page

The login page is the page that appears the first time the application is run. The login page is used as an intermediary to enter the application for users who have been granted access rights.
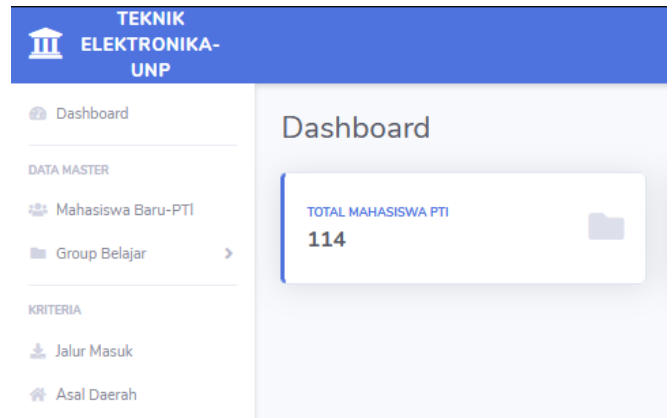
Dashboard Page

Figure 3. Dashboard Page

The admin dashboard page is the first page that appears after the admin logs in to the application

Data Student Page

Figure 4. Data Student Page

The student data page is a page that displays data for all PTI students. The student data displayed are name, number, gender, entry route, GPA, regional origin, school origin, cellphone/wam number, email, parent's occupation, economic level.

Data Import Page

Figure 5. Data Import Page

The student data import page is a page used by the admin user to import PTI student data from the Department of Electronics Engineering into an application with an .xlsx file format.

Update Data Page

Figure 6. Update Data Page

The student data update page is a page that contains a form used by the admin user to update student data if there is a change in PTI student data.

K-Means Clustering Analysis

Figure 7. K-Means Clustering Analysis

The k-means clustering analysis page displays data on all PTI students who are actively studying which will be processed using the k-means clustering algorithm based on the criteria and the value of each predetermined criterion. The following is a graph of the results of the clustering:

ExampleSet

Figure 8. Clustering Chart

The picture above informs the results of the clustering of this application providing information that members of cluster_0 are 33 students, cluster_1 is 28 students and cluster_2 is 53 students.

Grouping Results Page

Figure 9. Grouping Resault Page

The results page for grouping student learning classes on the admin is a page that displays the results of grouping student learning classes after the application clusters PTI student data.
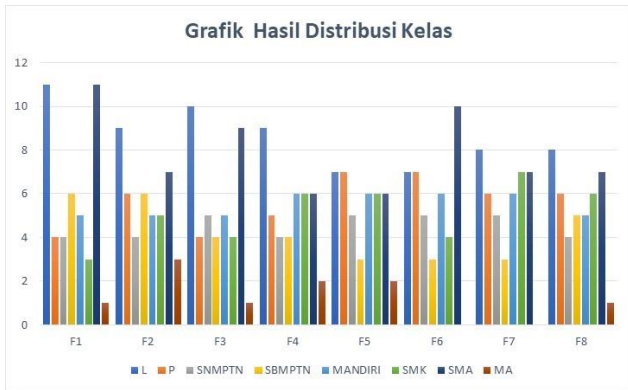
Figure 10. Grouping Resault Chart

Based on the graphic above, the results of the distribution of the study class for PTI Study Program students, Department of Electronics Engineering, FT-UNP can be described as follows:

Table 2. Grouping Resault

	F1	F2	F3	F4	F5	F6	F7	F8
L	11	9	10	9	7	7	8	8
P	4	6	4	5	7	7	6	6
SNMPTN	4	4	5	4	5	5	5	4
SBMPTN	6	6	4	4	3	3	3	5
MANDIRI	5	5	5	6	6	6	6	5
SMK	3	5	4	6	6	4	7	6
SMA	11	7	9	6	6	10	7	7
MA	1	3	1	2	2	0	0	1

Based on the table above, the application for compiling study groups for PTI FT-UNP Study Program students has arranged student study groups by grouping heterogeneity between study group members, and homogeneity between study groups.

CONCLUSION

Provide a statement that what is expected, as stated in the "INTRODUCTION" section can ultimately result in "RESULTS AND DISCUSSION" section, so there is compatibility. Moreover, it can also be added the prospect of the development of research results and application prospects of further studies into the next (based on result and discussion).

The results of the K-Means Clustering calculation are in the form of grouping study classes for PTI FT-UNP Study Program students which have almost the same student data and then divided into study classes so as to produce study class groupings, namely:

a. Each class has a variety of college entrance paths, namely the SNMPTN, SBMPTN and Mandiri pathways. Each class has a variety of college entrance paths, namely the SNMPTN, SBMPTN and Mandiri pathways.
b. Each class has a variation of GPA consisting of the lowest GPA to the highest GPA.
c. Each class has a ratio of the number of female (P) and male (L) sexes between classes.
d. Each class has a variety of school origins from SMA, SMK and MA.
e. Each class has a variation of regional origin consisting of within and outside the province of West Sumatra.

Each class has a variation of the level of economic ability which consists of a low level of economic ability to a low level of economic ability[16][17].

SUGGESTION

The following suggestions may be needed in the development of future student class clustering applications using the k-means clustering algorithm:

1. This application can be added features to make changes to study classes by the Management of the Department for certain reasons.
2. This application can be added to the number of clusters automatically.
3. This application can be added to the initial search for cluster centers so that the calculation results are more optimal.
4. This application can be added with supporting features for further development.

REFERENCES

[1] Agus Nur Khormarudin. (2016). Teknik Data Mining: Algoritma K-Means Clustering. *Jurnal Ilmu Komputer*.

[2] Asriningtias, S., & Sonalitha, E. (2018). Aplikasi k-means clustering untuk pengelompokkan siswa ke dalam kelas berdasarkan nilai akademis, jenis kelamin, perilaku dan nama siswa di sma negeri 1 srengat. *Seminar Nasional Sistem Informasi*, 1179–1187.

[3] Dr. Wuri Wuryandani, S.Pd., M. P. (2014). *Interaksi Sosial dan Ability Grouping*. Staffnew.Uny.Ac.Id.http://staffnew.uny.ac.id/upload/132309073/pengabdian/sapen18-

november-2014.pdf

[4] Firidho, M. A. (2019). *Studi Komparasi Penerapan Kelas Homogen Dan Heterogen Terhadap Motivasi Belajar di MA Al-Fatich Suarabaya*. Universitas Islam Negeri Sunan Ampel.

[5] Fauzia, S., Eka, F., Agustin, M., Syaripudin, U., & Ichsani, Y. (2016). Perancangan Prototype Tampilan Antarmuka Pengguna Aplikasi Web Kamardagang . Com Dengan Teknik Flat Design Pada Pt . *Teknik Informatika*, *9*(2), 148–157.

[6] Harli, E., Fauzi, A., & Kusmanto, T. H. (2016). Pengelompokkan Kelas Menggunakan Self Organizing Map Neural Network pada SMK N 1 Depok. *Jurnal Edukasi Dan Penelitian Informatika (JEPIN)*, *2*(2), 90–95. https://doi.org/10.26418/jp.v2i2.17574

[7] Iqbal, A. F. (2018). *Efektivitas kelas belajar homogen dalam pembentukan karakter peserta didik di SMP Bilingual Terpadu Krian Sidoarjo*. 2018. http://digilib.uinsby.ac.id/id/eprint/24645

[8] Kadar, S. & A. (2019). *Tingkat Kemampuan Awal Mahasiswa Prodi Pendidikan Matematika Fkip Universitas Dr.Soetomo Surabaya Ditinjau Dari Asal Daerah*.

[9] Muhammad Yusril Helmi Setyawan, A. S. M. (2020). *Panduan Lengkap Membangun Sistem Monitoring Kinerja Mahasiswa Internship Berbasis Web Dan Global Positioning System*. Kreatif Industri Nusantara. https://books.google.co.id/books?id=wqLzDw AAQBAJ

[10] Pasina, I., Bayram, G., Labib, W., Abdelhadi, A., & Nurunnabi, M. (2019). Clustering students into groups according to their learning style. *MethodsX*, *6*(March), 2189–2197. https://doi.org/10.1016/j.mex.2019.09.026

[11] Sari, R. R. (2018). Pengaruh Latar Belakang Sosial Ekonomi Keluarga Dan Self-Efficacy Terhadap Hasil Belajar Siswa Pada Mata Pelajaran Ekonomi (Survei Pada Siswa Kelas Xi Iis Sma Negeri Se-Kabupaten Subang Wilayah Selatan Tahun Ajaran 2017-2018). Universitas Pasundan.

[12] Winardi, A. (2016). *Artikel Skripsi Universitas Nusantara PGRI Kediri ANALISIS PENERAPAN METODE*. Ekonomi Akuntansi, 01(08), 1–13.

[13] Sianturi, F. A., Hasugian, P. M., Simangunsong, A., Nadeak, B., & Sihotang, H. T. (2019). *DATA MINING: Teori dan Aplikasi Weka*. IOCS Publisher. https://books.google.co.id/books?id=MWcHE

[14] Sianturi, F. A., Hasugian, P. M., Simangunsong, A., Nadeak, B., & Sihotang, H. T. (2019). *DATA MINING: Teori dan Aplikasi Weka*. IOCS Publisher. https://books.google.co.id/books?id=MWcHE AAAQBAJ

[15] Wuryandani, Wuri, *Interaksi Sosial dan Ability Grouping*.Website:http://staffnew.uny.ac.id/upload/132309073/pengabdian/sapen18-november-2014.pdf, diakses tanggal 25 Oktober 2021.

[16] F. F. Ranuharja, I. P. Dewi, others, G. Ambiyar, A. D. S. Fadhilah, and Y. Indarta, "Feasibility Test Assessment of LENTERA as Learning Source Media," *Res. Tech. Vocat. Educ. Train.*, vol. 1, no. 1, pp. 41–49, 2022.