EKSAKTA
Journal of Sciences and Data Analysis

E-ISSN: 2720-9326
P-ISSN: 2716-0459
DOI: 10.20885/EKSAKTA.vol3.iss2.art7

Research Article

Clustering of PDQ Participant Student in Faculty of Mathematics and Natural Sciences UII using the ROCK Method

Rahmadi Yotenka[1], Sekti Kartika Dini[1*]

[1] Department of Statistics, Faculty of Mathematics and Natural Sciences, Universitas Islam Indonesia

[*] *Corresponding author: sektidini@uii.ac.id*

Received: 2 June 2022; Accepted: 1 September 2022; Published: 7 September 2022

Abstract: The Qur'anic Self-Development (PDQ)-Ta'lim Program is one of the student activities that must be followed by diploma and bachelor program students in Universitas Islam Indonesia (UII). The implementation of PDQ is coordinated by each faculty which is carried out for 4 semesters with 12 meetings for each semester. After carrying out PDQ activities, it is necessary to know the student profiles that can be used as the basis for policy making in the implementation of PDQ activities in the next period. In order to find out the profile of students after participating in PDQ activities, it is necessary to group these students based on related variables. This study uses the ROCK method to group students participating in the PDQ Faculty of Mathematics and Natural Sciences (FMIPA) UII batch 2020. The ROCK method is a robust agglomerative hierarchical-clustering algorithm based on the notion of links. The ROCK method is a suitable clustering method for grouping data with categorical variables. Based on the results of the analysis of the ROCK method of student data for the batch 2020 FMIPA UII, obtained three optimum clusters (k=3) at a threshold value of θ of 0.20. Threshold 0.20 has the smallest S_W/S_B ratio value of 0.0514 or 5.14% and the largest R-squared value is 61.76% compared to other thresholds.

Keywords: Clustering, Categorical Variables, PDQ Participants, ROCK.

Introduction

Universitas Islam Indonesia (UII) as one of the Islamic higher education institutions in Indonesia, wants to produce graduates who can develop, implement, and practice Islamic values in society. The implementation of education based on Islamic values contains standard values of Islamic development that are embedded in student activities with the aim of forming character and producing educated graduates with Islamic morals [1].

Based on UII Rector Regulation No. 24 of 2019 concerning the Participation and Activity Credit Units for the Bachelor level which are mandatory, including: The Deepening of Basic Islamic Values (PNDI), The Qur'anic Self-Development (PDQ), Leadership Training and Da'wah (PKD), and Self Development Training (PPD). These activities are appreciated in the form of Participation Credit Unit (SKP) with a total weight of 50 SKP [2].

The Qur'anic Self-Development (PDQ)-Ta'lim program is one of the student activities that must be followed by students of diploma and bachelor programs at the UII. The PDQ study material focuses on reading and writing the Qur'an, memorizing the Qur'an juz 30, memorizing selected hadith, and learning Arabic to understand the Qur'an. The implementation of PDQ is coordinated by each faculty with a load of 20 SKP. PDQ activities run for 4 semesters with 12 meetings for each semester [3].

The series of PDQ activities begins with Placement Test Agama (PTA) for new students. The results from the PTA will be used to group students into four levels, namely Pre-basic, Elementary, Intermediate, and Advanced levels. Each group has PDQ achievements and targets as outlined in Student

Activity Lesson Plan (RPAK) [1]. At the end of each semester there is a PDQ exam for each level with the aim of evaluating student learning outcomes.

After carrying out the PDQ activity, it is necessary to know the profile or characteristics of students. This needs to be done as a basis for policy making in the implementation of PDQ activities in the next period. In order to know the profile of students after participating in PDQ activities, it is necessary to group these students based on related variables.

Cluster analysis is a multivariate analysis that aims to classify objects into several groups based on their characteristics. In selecting the method for cluster analysis, it is important to consider the data type of the variables that form the basis for grouping objects. This relates to the similarity measure between the objects used.

In order to group the PDQ participant students' FMIPA UII with the variables used were categorical variables, such as gender, study program, GPA, high school origin of the student, reading score, memorization score, worship score, and prayer score, an appropriate cluster analysis method are needed. The traditional hierarchical method is considered inappropriate for grouping categorical data. So the ROCK (Robust Clustering using links) method was developed to group data with these categorical variables.

ROCK is a robust agglomerative hierarchical-clustering algorithm based on the notion of links. The ROCK algorithm is an appropriate algorithm for grouping data with categorical variables. To merge data points, ROCK uses links between data points instead of distances between data. Link is used to measuring the similarity/closeness between a pair of data points. Observations that have a high link level are combined into one group, while those with a small link value are separated from the group where the data is grouped. In some cases, the ROCK method can also handle outlier data. Outlier parts of the data can appear as a group [4].

Based on the explanation above, this research uses the ROCK method which aims to group the PDQ participant students' FMIPA UII batch 2020 whose results are expected to provide an overview of the characteristics of PDQ participants who have attended the coaching. In addition, it is hoped that the results of this study can be used as input for the student affairs field for student development activities in the Islamic field.

Materials and Methods

Data

The data used in this study is sourced from UII student academic data (UII Gateway), including gender, study program, GPA, and high school origin of the student [5]. Meanwhile, data on reading scores, memorization scores, worship scores, and prayer scores were obtained from the Director of Islamic Development and Education (DPPAI) UII data recap. The data description for each variable is shown in Table 1.

Table 1. PDQ FMIPA categorical data variables

Number	Variables	Category	Explanation
1	Study Program	1	Statistics
		2	Chemistry
		3	Pharmacy
		4	Chemistry Education
		5	Chemical Analysis
2	School	1	Public High School
		2	Private High School
		3	Public Vocational High School
		4	Private Vocational School
		5	Islamic High School

3	GPA	1	Summa Cumlaude
		2	Cumlaude
		3	Very satisfactory
		4	Satisfactory
		5	Good enough
		6	Less satisfactory
4	Reading	1	Very Poor
		2	Poor
		3	Good
		4	Very Good
5	Memorization	1	Very Poor
		2	Poor
		3	Good
		4	Very Good
6	Worship	1	Very Poor
		2	Poor
		3	Good
		4	Very Good
7	Prayer	1	Very Poor
		2	Poor
		3	Good
		4	Very Good
8	Gender	1	Male
		2	Female

Data Analysis

Descriptive Statistics

Descriptive Statistics is part of the Statistical method on how to collect numbers in the form of notes and then how to present these numbers in the form of graphs and tables [6]. A graph or diagram is a visual presentation of data in the form of images. Meanwhile, a table is a presentation of data arranged according to certain categories in a list [7].

The main characteristic of a graphic is that it has a simple visual appearance, but its content includes all the information and is easy to understand. One type of graph that is often used to present categorical type data is a bar graph. Bar graphs are used to compare values across categories using vertical bars [8].

Cluster Analysis

Cluster analysis is an analytical method that aims to group a set of objects into a number of groups based on the similarity measure between the objects [9]. Cluster analysis consists of two types, namely hierarchical and non-hierarchical. The basic difference between the two types of cluster analysis is in the formation of the number of groups.

Hierarchical cluster analysis is a cluster analysis method that seeks to build a group hierarchy [10]. The strategy for forming groups in hierarchical cluster analysis is generally divided into two, namely agglomerative and divisive. Agglomerative processing, where each object is considered as a separate group, then two groups that have similarities are combined into a new group, and so the next step. Meanwhile, the divisive process starts from a large group consisting of all objects, then the objects with the highest differences are separated from the large group and so on. In the non-hierarchical cluster analysis, it begins by determining the number of groups to be formed, then the grouping process can be carried out without following the process that occurs in the hierarchical method [11].

Robust Clustering using linKs (ROCK)

The ROCK method was developed from the agglomerative hierarchical clustering method used for categorical data by using a link to measure the similarity between a pair of points [4] [12]. Objects with high links (relationship level) will be combined in one group, while objects with low link (relationship level) will be separated from the group. The following are the steps in the ROCK method [4]:

a. Initialize objects as a group with one member
b. Calculating the similarity between groups using the Jaccard coefficient (Equation 1)

$$sim(X_i, X_j) = \frac{|X_i \cap X_j|}{|X_i \cup X_j|}, \qquad i \neq j \tag{1}$$

where X_i is the categorical set of the *i*-th observations, while X_j is the categorical set of *j*-th observations. A categorical set is a set whose members are categorical data. $|X_i \cap X_j|$ represents the number of identical members between X_i and X_j, and $|X_i \cup X_j|$ represents the number of joint members X_i and X_j.

c. Determine the number of clusters (n) and determine the threshold value (θ) to form a neighboring matrix. The value of θ is the parameter chosen by the researcher to determine the close relationship between observations, whether a couple of observations are neighbors or not. The threshold value of θ between 0 and 1.
d. Form a neighbor matrix **A** based on the value of θ. The element of neighbor matrix **A** in the i-th row, j-th column is 1 if $sim(X_i, X_j) \geq \theta$ and 0 if $sim(X_i, X_j) < \theta$.
e. Count the number of links, there are two ways that can be explained as follows:
   i. The link between two clusters with each cluster contains 1 observation:
   If C_i is the i-th cluster that has one observation X_i then the link is calculated by Equation 2.

$$Link\ (C_i, C_j) = Link(X_i, X_j) = |T_{X_i} \cap T_{X_j}|, i \neq j \tag{2}$$

   where T_{X_i} is the neighboring set of the i-th observation, T_{X_j} is the neighboring set of the j observations, and $|T_{X_i} \cap T_{X_j}|$ is the number of same elements of the neighboring set between X_i and X_j.
   ii. The link between two clusters with each cluster contains more than 1 observation:
   If C_i is the i-th cluster which is the set of X_i members that are placed into one group, then the link is calculated by Equation 3.

$$Link\ (C_i, C_j) = \sum_{X_i \in C_i,\, X_j \in C_j} Link\ (X_i, X_j), i \neq j \tag{3}$$

   where $Link\ (C_i, C_j)$ represents the number of links between each possible pair of observations that are in C_i and C_j
f. Calculate the value of the goodness measure to join two groups using Equation 4

$$g(C_i, C_j) = \frac{Link(C_i, C_j)}{(n_i + n_j)^{1+2f(\theta)} - {n_i}^{1+2f(\theta)} - {n_j}^{1+2f(\theta)}} \tag{4}$$

where n_i and n_j each represent the number of members in the i-th group and j-th group and $f(\theta) = \frac{1-\theta}{1+\theta}$

g. Merge groups that have the largest goodness measure, then recalculate the links between groups and update the new goodness measure value.
h. Perform steps f and g until the expected number of groups is reached or until there are no more links between groups
i. Repeat steps a to h with different values of θ
j. Calculate the ratio between S_W and S_B for each values of θ by using Equation 5 and Equation 6
k. Compare the results of step j for each value of θ to determine the optimum value of θ based on the criteria of a small ratio of S_W and S_B

E-ISSN: 2720-9326
P-ISSN: 2716-0459

ROCK Clustering Validation

An important step in cluster analysis is the determination of the optimum number of clusters and the optimum threshold value (θ). Clustering validation is needed to measure the performance of clustering results [12]. Clustering validation is a quantitative and objective evaluation process of cluster analysis results [13]. A good clustering process will produce clusters with high homogeneity between members in the same cluster and high heterogeneity with members of different clusters [14].

Clustering performance for numerical data can be determined based on the ratio of the standard deviation values in within group (S_W) and the standard deviation between groups (S_B). The smaller the value of the ratio (S_W) and (S_B) indicates the better the performance of the clustering method used [15]. Performance measurement using the ratio (S_W) and (S_B) is only used for numerical data, meanwhile for categorical data using a contingency table which is equivalent to ANOVA (Analysis of Variance). The measure of diversity for categorical data based on the standard deviation in within group (S_W) and the standard deviation between groups (S_B) can be seen in Equation 5 and Equation 6 [16] [17].

$$S_W = \left[\frac{WSS}{n-c}\right]^{1/2} \tag{5}$$

$$S_B = \left[\frac{BSS}{c-1}\right]^{1/2} \tag{6}$$

where $WSS = \frac{n}{2} - \frac{1}{2}\sum_{c=1}^{C}\frac{1}{n_c}\sum_{k=1}^{K} n_{kc}^2$ dan $BSS = \frac{1}{2}\left(\sum_{c=1}^{C}\frac{1}{n_c}\sum_{k=1}^{K} n_{kc}^2\right) - \frac{1}{2n}\sum_{k=1}^{K} n_k^2$

where if there are n observations with n_k is the number of observations with the k-th category, where k = *1,2,3,…,K* and $\sum_{k=1}^{K} n_k = n$. Then n_{kc} is the number of observations with the k-th category and the c-th cluster, where c = *1,2,3,…,C* and C is the number of clusters formed.

Results and Discussions

Descriptive Statistics

Based on UII student data at https://gateway.uii.ac.id/akademik, the number of students of FMIPA UII batch 2020 for Diploma and Bachelor's level who are active from odd semester 2020/2021 to odd semester 2021/2022 are 471 students. Table 1 shows that the majority of FMIPA UII students in the batch 2020 are female (80.3%) and majority are graduates of public high school (58.4%) and Islamic high school (26.8%). To continue their higher education at UII, 35.9% chose Pharmacy Study Program, 35% Statistics Study Program, 13.2% Chemistry Study Program, 11.3% Chemical Analysis Study Program, and 4.7% chose Chemistry Education Study Program.

Table 1 also provides information on the results of the Religion Placement Test (PTA) sourced from the DPPAI UII. The ability to read the Qur'an (Reading variable) and memorize the Qur'an Juz 30 (Memorization variable) the majority of the students are "good" with the respective percentages being 47.3% and 40.6%. This is directly proportional to the last GPA percentage obtained by the majority of students, which is 72.4% in the Cumlaude category. On the other hand, the knowledge about the practice of Islamic religious worship (Worship variable) and the ability to memorize daily prayers (Prayer variable) of the students were mostly in the "poor" category, respectively at 46.7% and 40.8%. The proportion of Reading, Memorizing, Worship, and Prayer variables for each Study Program based on the gender of students is shown in Figure 1. to Figure 4.

E-ISSN: 2720-9326
P-ISSN: 2716-0459

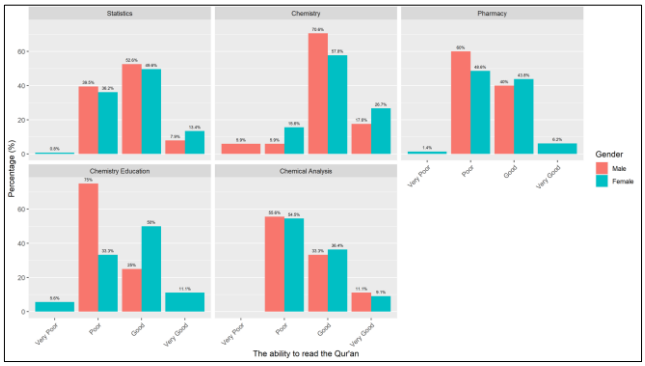

Figure 1. Proportion of reading the Qur'an for each study program based on gender

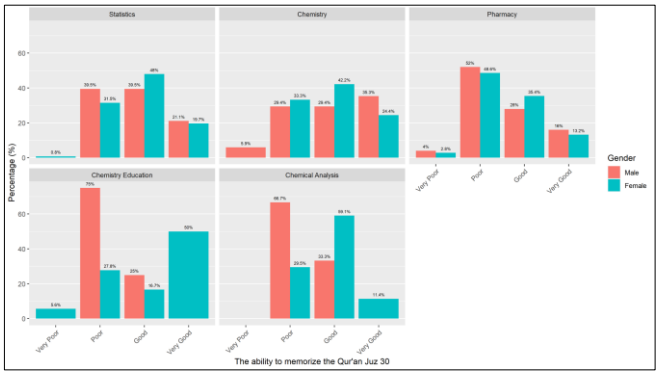

Figure 2. Proportion of memorizing the Qur'an juz 30 of each study program based on gender

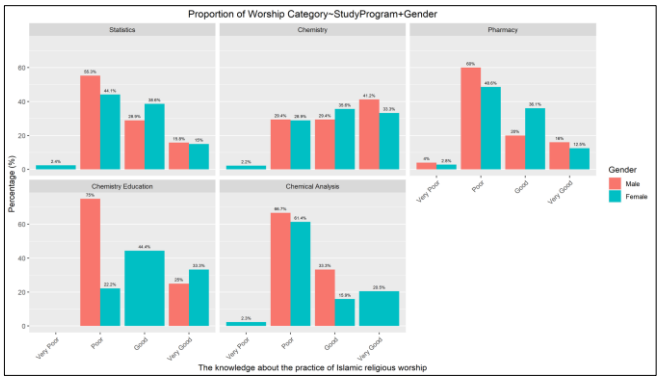

Figure 3. The proportion of understanding of the practice of Islamic religious worship for each study program is based on gender

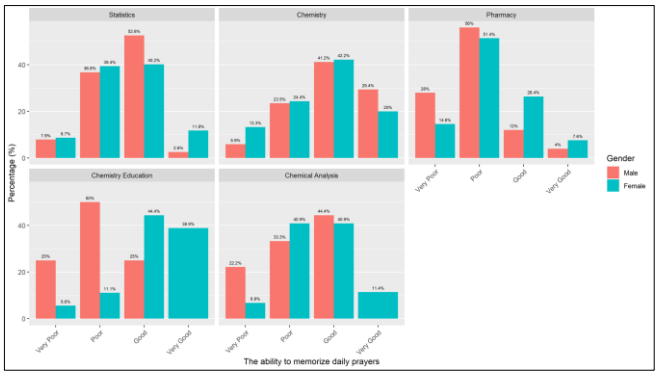

Figure 4. Proportion of memorizing daily prayers for each study program based on gender

Figure 1. to Figure 4. shows that the majority of reading the Qur'an, memorizing the Qur'an juz 30, and memorizing the daily prayers of male and female students of Statistics Study Program are categorized as "good". However, the understanding of the practice of Islamic religious worship is mostly in the "poor" category. In the Chemistry Study Program, the majority of male and female students are categorized as "good" for reading the Qur'an, memorizing the Qur'an juz 30, understanding the practice of Islamic worship, and memorizing daily prayers. On the other hand, the majority of reading the Qur'an, memorizing the Qur'an juz 30, understanding the practice of Islamic worship, and memorizing the daily prayers of male and female students of Pharmacy Study Program are categorized as "poor".

In the Chemical Analysis Study Program, the majority of reading the Qur'an, memorizing the Qur'an juz 30, and understanding the practice of Islamic religious worship, especially male students are categorized as "poor". However, most of the students' memorization of daily prayers, both male and female, was in the "good" category. Then in the Chemistry Education Study Program, the majority of male students were categorized as "poor" for reading the Qur'an, memorizing the Qur'an juz 30, understanding the practice of Islamic worship, and memorizing daily prayers. On the other hand, the majority of female students of Chemistry Education Study Program are categorized as "good" for reading the Qur'an, memorizing the Qur'an juz 30, understanding the practice of Islamic worship, and memorizing daily prayers.

Table 2. Description of the categorical variables of FMIPA UII students in the class of 2020

	Percentage Category Variable (%)						
Variables	Category "1"	Category "2"	Category "3"	Category "4"	Category "5"	Category "6"	Total
Study Program	35.0	13.2	35.9	4.7	11.3	0.0	100
School	58.4	4.0	2.3	8.5	26.8	0.0	100
GPA	1.1	72.4	20.6	2.5	2.1	1.3	100
Reading	1.1	40.8	47.3	10.8	0.0	0.0	100
Memorization	1.7	39.3	40.6	18.5	0.0	0.0	100
Worship	2.1	46.7	33.1	18.0	0.0	0.0	100
Prayer	11.9	40.8	35.9	11.5	0.0	0.0	100
Gender	19.7	80.3	0.0	0.0	0.0	0.0	100

The Results of ROCK Clustering

The first step in clustering is to assume each object as a cluster with a single member. Then calculate the distance matrix (sim) measuring 471 x 471 to detect the similarity (similarity) between objects.

$$sim = \begin{bmatrix} 1 & 0.78 & 0.33 & 0.23 & 0.23 & \cdots & 0.45 & 0.45 & 0.23 & 0.23 & 0.14 \\ 0.78 & 1 & 0.33 & 0.14 & 0.14 & \cdots & 0.60 & 0.33 & 0.14 & 0.14 & 0.07 \\ 0.33 & 0.33 & 1 & 0.14 & 0.33 & \cdots & 0.33 & 0.33 & 0.14 & 0.14 & 0.14 \\ 0.23 & 0.14 & 0.14 & 1 & 0.14 & \cdots & 0.07 & 0.23 & 0.14 & 0.23 & 0.07 \\ 0.23 & 0.14 & 0.33 & 0.14 & 1 & \cdots & 0.14 & 0.23 & 0.23 & 0.23 & 0.45 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0.45 & 0.60 & 0.33 & 0.07 & 0.14 & \cdots & 1 & 0.60 & 0.23 & 0.33 & 0.23 \\ 0.45 & 0.33 & 0.33 & 0.23 & 0.23 & \cdots & 0.60 & 1 & 0.33 & 0.60 & 0.33 \\ 0.23 & 0.14 & 0.14 & 0.14 & 0.23 & \cdots & 0.23 & 0.33 & 1 & 0.23 & 0.60 \\ 0.23 & 0.14 & 0.14 & 0.23 & 0.23 & \cdots & 0.33 & 0.60 & 0.23 & 1 & 0.45 \\ 0.14 & 0.07 & 0.14 & 0.07 & 0.45 & \cdots & 0.23 & 0.33 & 0.60 & 0.45 & 1 \end{bmatrix}$$

E-ISSN: 2720-9326
P-ISSN: 2716-0459

Figure 5. Sim matrix visualized in the form of a heatmap.

The heatmap shows the process of grouping research objects with graded colors. The colors in the heatmap show the similarity between objects.

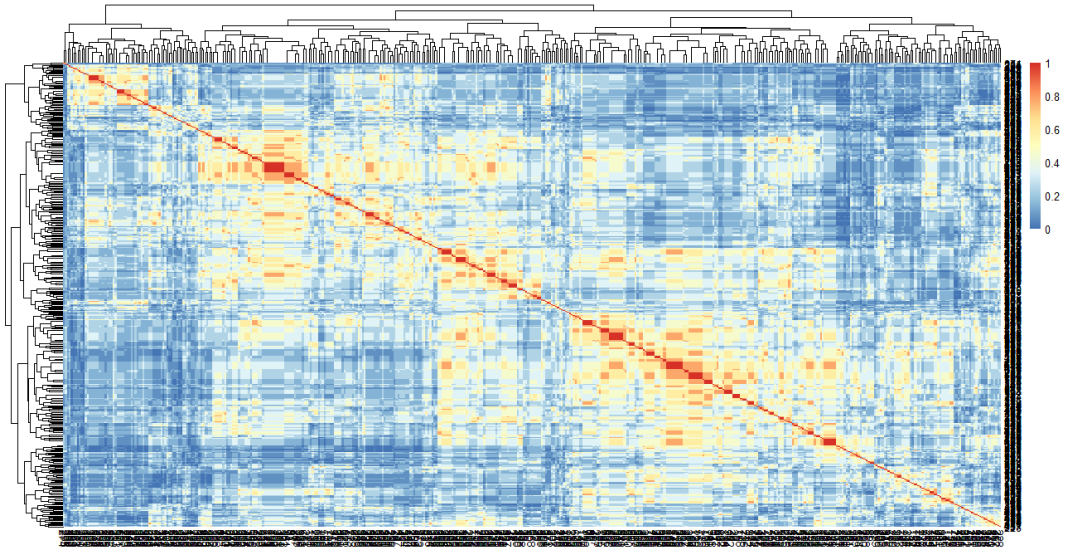

Figure 6. Heatmap sim matrix size 471 *x* 471

After detecting the similarity of the object with the distance matrix, the next step in ROCK clustering is to determine the threshold value (θ), calculate the ratio S_W/S_B, and calculate the R-squared used to form the optimal number of clusters. Threshold value (θ) is the limit for determining neighbors which is then used to calculate the link matrix and goodness measure. An object is said to be neighbor to another object if the sim value > θ. In this study, the θ is 0.01, 0.07, 0.10, 0.15, 0.20, 0.25, 0.27, and 0.30. The best number of clusters (optimal) is determined from the smallest S_W/S_B ratio value at one of the thresholds value. The results of the calculation of the S_W/S_B and R-squared for each threshold are shown in Table 3.

Table 3. S_W/S_B Ratio and R-squared based on the number of clusters

Threshold value (θ)	Number of *cluster*	S_W/S_B Ratio	R-squared
0.01	2	0.3692	0.0010
0.07	3	0.1100	0.2611
0.10	3	0.0803	0.3985
0.15	3	0.1374	0.1846
0.20	3	0.0514	0.6176
0.25	4	0.3237	0.0578
0.27	2	0.0725	0.2883
0.30	2	0.0726	0.2881

Threshold value 0.20 has the smallest S_W/S_B ratio value of 0.0514, as summarized in Table 3. The R-squared value obtained is also the highest compared to other thresholds, namely 0.6176 or 61.76%. Thus, the best number of clusters (optimal) with the ROCK method using a threshold value (θ) 0.20 is 3

E-ISSN: 2720-9326
P-ISSN: 2716-0459

clusters (Figure 7). Cluster 1 is indicated by the color pink with the number of cluster members as many as 238 students, cluster 2 is marked in red with the number of cluster members as many as 23 students, and cluster 3 is indicated by the blue color with the number of cluster members as many as 210 students. Figure 8. shows the distribution of students for each cluster based on the study program.

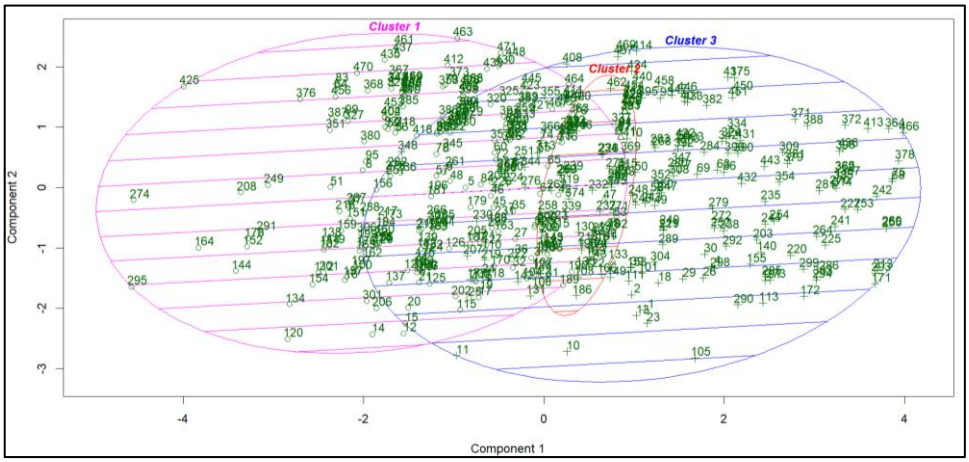

Figure 7. Distribution of students per cluster

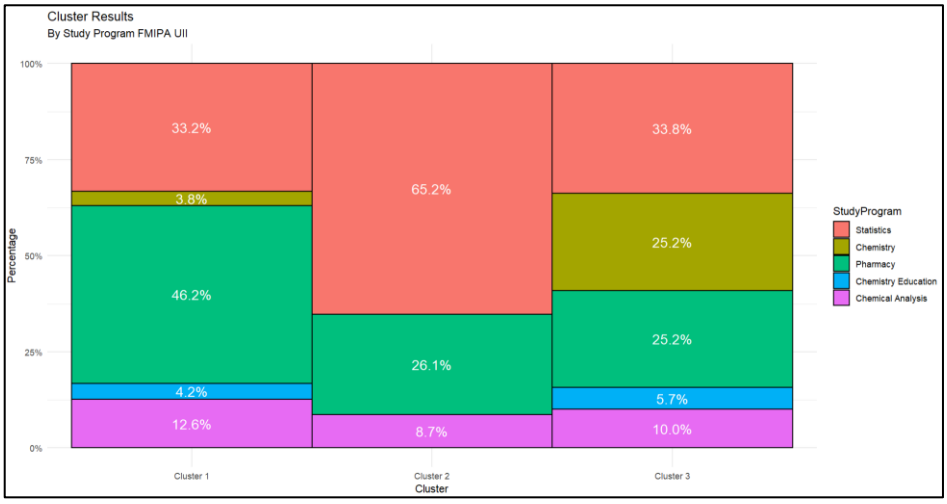

Figure 8. Clustering of students by study program

Based on the barplot of Figure 8., the members of cluster 1 and cluster 2 are dominated by students from the Pharmacy Study Program and Statistics Study Program. Cluster 3 members is dominated by students from Statistics Study Program, Pharmacy Study Program, and Chemistry Study Program.

Then the summary of characteristics or profiles of each cluster resulting from the ROCK clustering analysis is shown in Table 4. The profile of each cluster is also compared with the profile of the object, namely all students of FMIPA UII batch 2020.

Table 4. Comparison of the profiling of each cluster with all objects

Variables	All Students	*Cluster* 1	*Cluster* 2	*Cluster* 3
Study Program	1	3	1	1
School	1	1	1	5
GPA	2	2	2	2
Reading	3	2	3	3
Memorization	3	2	3	3

E-ISSN: 2720-9326
P-ISSN: 2716-0459

Variables	All Students	*Cluster* 1	*Cluster* 2	*Cluster* 3
Worship	2	2	3	4
Prayer	2	2	3	3
Gender	2	2	2	2

Based on Table 4., cluster 1, cluster 2, and cluster 3 there is no difference in characteristics for the "GPA" variable and the "Gender" variable. This means that the three clusters are mostly female student groups who have a Cumlaude GPA. This fact is shown in more detail in Figure 9.

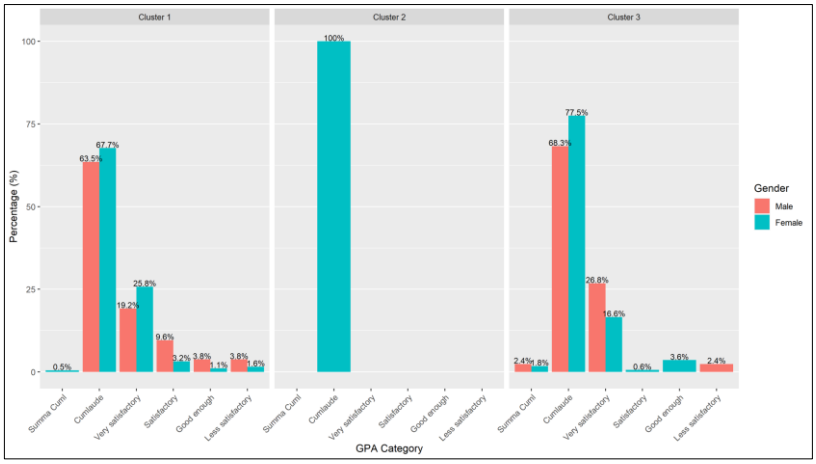

Figure 9. Student clustering based on GPA and gender

For the variables "Study Program", "School", "Reading", "Memorization", "Worship", and "Prayer" have different characteristics between cluster 1, cluster 2, and cluster 3. Overall, the segmentation of the PDQ participant students' FMIPA UII batch 2020 with the eight characterizing variables can be summarized in Table 5.

Table 5. Comparison of the profiling for each cluster with all objects

Cluster	Characteristics of the PDQ participant students' FMIPA UII batch 2020
1	The majority are students from the Pharmacy Study Program, from public high schools, have a Cumlaude GPA, the ability of read the Qur'an in the "poot" category (there are some errors in reading the Qur'an, especially in terms of recitation and makharijul letters), the ability to memorize Al-Qur'an The Qur'an juz 30 is in the "poor" category, the understanding of Islamic religious practices is in the "poor" category, the ability to memorize daily prayers is in the "poor" category, and the majority are female.
2	The majority are Statistics Study Program students, coming from public high schools, have a Cumlaude GPA, the ability to read the Qur'an in the "good" category (fluently read the Qur'an with errors in tajwid and makharijul letters relatively no more than one error every one verse of Al-Qur'an), the ability to memorize Al-Qur'an juz 30 is in the "good" category,

Cluster	Characteristics of the PDQ participant students' FMIPA UII batch 2020
3	understanding the practice of Islamic religious worship is in the "good" category, the ability to memorize daily prayers is in the "good" category, and the majority are female. The majority are Statistics Study Program students, come from islamic high school, have a Cumlaude GPA, the ability to read the Qur'an in the "good" category (fluently read the Qur'an with errors in tajwid and makharijul letters relatively no more than one error every one verse of Al-Qur'an), the ability to memorize Al-Qur'an juz 30 is in the "good" category, the understanding of Islamic religious practices is in the "very good" category, the ability to memorize daily prayers is in the "good" category, and the majority are female.

If we look at the dominance of categories on the variables "Reading", "Memorization", "Worship", and "Prayer" in each cluster, it can be said that the first cluster is a poor cluster. This means that the students of FMIPA UII batch 2020 group in the first cluster has poor potential in understanding Islamic religious knowledge, especially from the ability to read the Qur'an, memorize Al-Qur'an juz 30, understand the practice of Islamic worship, and memorize prayers daily. While the second and third clusters are groups of students who have good potential in understanding Islamic religious knowledge.

More detail at the "GPA" variable category as shown in Figure 9., it can be said that the first cluster is a group of students who have the potential to graduate late because the percentage of students with a GPA $\leq$ 3.00 (some have a GPA < 2.25) is 8.4%. The third cluster of students also has the potential for late graduation because there are still students with a GPA $\leq$ 3.00 with a percentage of around 4%. Meanwhile, the second cluster is a group of students with the most potential for faster graduation times than the first and third clusters because all students in this cluster have a cumlaude GPA (3.51-3.99). Overall, it can be said that cluster 2 is the best group of students in terms of academics and understanding of Islamic religious knowledge. Meanwhile, cluster 1 is a group of students who most need special attention and treatment to improve academic potential and understanding of Islamic religious knowledge, especially the ability to read the Qur'an, memorize Al-Qur'an juz 30, understand the practice of Islamic worship, and memorize daily prayer.

Conclusion

Based on the results of the ROCK clustering analysis of PDQ participant students' FMIPA UII batch 2020, obtained 3 optimum clusters (k=3) at a threshold value (θ) of 0.20. Threshold 0.20 has the smallest S_W/S_B ratio value of 0.0514 or 5.14% and the largest R-squared value is 61.76% compared to other thresholds. The number of members of cluster 1 is 238 students, members of cluster 2 are 23 students, and members of cluster 3 are 210 students. The three clusters are dominated by female student groups who have a Cumlaude GPA category.

Overall, cluster 2 is the best student cluster in terms of academics (the GPA of all cluster members is categorized as cumlaude) and also the understanding of Islamic religious knowledge. So that the student groups in cluster 2 can develop their potential both in the academic field and in the religious field. In the academic field, for example, it can be recommended to become a practicum assistant, lecturer assistant, delegate for scientific competitions, and collaborate on community service and research activities with lecturers. In the religious field, it can be recommended to be a supervisor for UII student religious activities such as PDQ-Ta'lim Program for students of the next batch.

Likewise with the student group in cluster 3, it can be recommended the same thing as cluster 2 because the majority of students have a cumlaude GPA (76%) and summa cumlaude (2%) and the majority of students have a very good potential for understanding Islamic religious knowledge. However, in cluster 3 also there are students with a GPA $\leq$ 3.00 (4%). So it requires special attention and treatment

from the related study program. This is related to the graduation time of the group of students who have the potential to pass more than the target.

In cluster 1, 8.4% of students also need special attention and treatment to increase their academic potential because they have a GPA ≤ 3.00. Then the majority of students in cluster 1, the understanding of Islamic religious science is categorized as poor. So that the management of religious activities, especially the DPPAI UII can provide more intense assistance with the right method for the students in cluster 1 to improve understanding of Islamic religious knowledge, especially the ability to read the Qur'an, memorize the Qur'an juz 30, understand the practice of religious worship. Islam, and memorizing daily prayers.

References

[1] DPPAI UII, Pedoman Pengembangan Diri Qur'ani (PDQ) Universitas Islam Indonesia Jenjang Sarjana. Yogyakarta: DPPAI UII, 2020.
[2] UII, Peraturan Rektor UII tentang SKP dan Aktivitas Kemahasiswaan. 2019.
[3] DPPAI UII, PDQ/TA'LIM, 2021. https://dppai.uii.ac.id/ondi/.
[4] S. Guha, R. Rastogi, and K. Shim, Rock: a robust clustering algorithm for categorical attributes, Inf. Syst., 25(5) (2000) 345–366.
[5] UII gateway. https://gateway.uii.ac.id/akademik.
[6] V. Silvia, Statistika Deskriptif, Penerbit Andi, Banda Aceh, 2020.
[7] L. M. Nasution, Statistik Deskriptif, J. Hikmah, 14 (1) (2017) 49–55.
[8] D. Setyowati, Pelatihan Membuat Grafik Dalam Microsoft Excel. Untuk Pengolahan Dan Penyajian Data, J. Dharma, 2(2) (2019) 101–112.
[9] J. Hair, R. Anderson, B. Babin, W. Black, Multivariate Data Analysis, 7th ed. Prentice Hall Inc, New Jersey, 2000.
[10] P.Baby and K.Sasirekha, Agglomerative Hierarchical Clustering Algorithm- A Review, Int. J. Sci. Res. Publ., 3(3) (2013) 2–4.
[11] S. Sulastri, L. Usman, and U. D. Syafitri, K-prototypes Algorithm for Clustering Schools Based on The Student Admission Data in IPB University, Indones. J. Stat. Its Appl., 5 (2) (2021) 228-242.
[12] M. A. Putri and S. Abdullah, Clustering analysis of senior high school in West Java based on educational facilities, J. Phys. Conf. Ser., 1725(1) (2021) 1-11
[13] A. K. Jain and R. C. Dubes, Clustering_Jain_Dubes.pdf. Prentice Hall Inc, New Jersey: 1988.
[14] J. Reddy and B. Kavitha, Clustering the Mixed Numerical and Categorical Datasets Using Similarity Weight and Filter Method, Int. J. Database Theory Appl., 5(1) (2012) 121–133.
[15] M. J. Bunkers and J. R. Miller Jr, Definition of Climate Regions in the Northern Plains Using an Objective Cluster Modification Technique, J. Clim., 9 (1996) 130–146.
[16] G. D. Kader and M. Perry, Variability for categorical variables, J. Stat. Educ., 15(2) (2007) 1-18.
[17] Alvionita, Sutikno, and A. Suharsono, Ensemble ROCK Methods and Ensemble SWFM Methods for Clustering of Cross Citrus Accessions Based on Mixed Numerical and Categorical Dataset, IOP. Conf. Ser. Earth Environ. Sci., 58 (2017) 1–11.