BAREKENG: Jurnal Ilmu Matematika dan Terapan
March 2022 Volume xx Issue xx Page xxx–xxx
P-ISSN: 1978-7227 E-ISSN: 2615-3017
https://doi.org/10.30598/barekengxxxxxxxxxxxxx

APPLICATION OF TWO GROUPS ANALYSIS AND CLUSTER ANALYSIS ON COMPARISON OF CHARACTERISTICS OF PAGARALAM COFFEE FARMERS CATEGORIES

Irmeilyana[1*], Ngudiantoro[2], Sri Indra Maiyanti[3]

[1,2,3]*Department of Mathematics, Faculty of Mathematics and Natural Science, University of Sriwijaya*
Jl. Raya Palembang - Prabumulih KM. 32, Indralaya, 30662, Indonesia

*Corresponding author e-mail: [1]*irmeilyana@unsri.ac.id*

Abstract. *Land care is one of the factors that must be considered in the sustainability of coffee production. A small number of Pagaralam coffee farmers know and apply reductant herbicide in weed control. This study aims to compare the characteristics of Pagaralam coffee farmers based on the use of reductant herbicide by using two groups analysis and cluster analysis. The characteristics of the farmers studied included 17 variables on 165 respondents who were selected by purposive sampling. The variables studied include land productivity and farmers' income. Respondents were divided into 2 categories, namely users and non-users of reductant herbicide. The initial stages of data processing in the form of descriptive statistics, hypothesis testing by mean difference test, correlation between variables, Principal Component Analysis, and biplot analysis. According to the results of the two groups analysis and cluster analysis, the variables dominantly tend to characterize the similarity in comparison between the categories of non-users and users are the variables that join into one cluster, namely land area, number of trees, coffee bean production, estimated yield, total harvest, Gross income, and Net income. While the variables that dominantly characterize the dissimilarity between the two categories are the variables that form 5 separate clusters, namely the clusters of plantation productivity, respondent identity, education, tree age, and length of harvest period.*

Keywords: *land productivity, Pagaralam coffee, cluster analysis, groups analysis, reductant herbicide.*

Article info:

Submitted: date, month, year *Accepted: date, month, year*

How to cite this article:
First author, second author, etc, "TITLE OF ARTICLE", *BAREKENG: J. Il. Mat. & Ter.*, vol. xx, iss. xx, pp. xxx-xxx, Month, Year.

1. INTRODUCTION

South Sumatra is a coffee-producing province with the highest production and land area in Indonesia, which respectively reached 25.39% and 20.09% of total production and area. This data was based on data from the Ministry of Agriculture in 2019. However, the productivity (in kg/ ha) ranks 4th after North Sumatra, Riau, and Jambi. At a fixed rate in 2019, South Sumatra's production, land area, and productivity decreased compared to 2018, which were 1.25%, 0.35%, and 1.06%, respectively [1].

Pagaralam robusta coffee which has had a GI (Geographical Indication) since 2020 have distinctive characteristics, including taste [2-3]. In the period 2016 to 2021 Pagaralam experienced a decline in coffee production. At [4], Pagaralam's coffee production reached 21,892 tons or 11.46% of coffee production in South Sumatra. While the estimation figures for 2021, Pagaralam's coffee production is 20,833 tons [1], or 11.04% of the estimated coffee production in South Sumatra.

Pagaralam coffee production is related to problems faced by farmers, including: characteristics and farming culture of farmers, maintenance costs factors (including capital to fertilize and control weeds), and marketing factors (including product selling prices) [5-9]. In addition, rainfall and temperature factors [10-11], lack of education to farmers about land care [12-15] and plant care, access to infrastructure, and the lack of education on post-harvest processing that has not been harmonized with access to marketing also contributed to the decline in coffee production.

The issue of sustainable agriculture is related to the mindset and culture of farming. According to [16] - [17], the concept is based on 3 aspects, namely: economic, social, and ecological. Sustainable agriculture can increase productivity and also farmers' income because organic products have a premium price. For example, organic rice production in DIY Province [18]. Its environmental impact on the process of improving soil fertility, rice production decreased significantly for 2 years. The sustainability of coffee production is highly dependent on land maintenance, which includes fertilization, the use of herbicides to control weeds, and pruning (rejuvenation). Weed control must be adapted to land conditions and weed types [19]. In [20], through education from several parties, coffee farmers in Rimba Candi Village feel the impact of the importance of using reductants in coffee fields.

The use of reductants can lead to sustainable agriculture. Reductants can reduce pesticide residues in agricultural areas and pesticide costs. Reductant mixtures in pesticides can save land maintenance costs [21-22]. Increasing the productivity of coffee farmers can be done by providing intensive training. Active coffee farmers in South Sumatra can be the focus of the training target [23]. Participatory extension methods such as “Farmer Field Schools” (FFS) can be used to increase awareness in sustainable land management [12].Reductant herbicide is introduced (by field workers from the manufacturer) to farmers through an educational process, so that coffee farmers can gain knowledge about land care and the use of reductant use can have a positive impact on the land of their users.

The mean difference test of 28 variables on 125 respondents of Pagaralam coffee farmers as users and non-users of reductants showed that only the average planting area per 1 tree, the age of the tree, the maximum selling price of coffee beans, and the number of workers were not the same between the two categories of respondents. The results of the independence test showed that there was a relationship between the respondent's category and the categories of each variable in the education, the frequency of herbicide use, and the farmers' perceptions of coffee production and income [23]. Research on the comparison of the characteristics of the categorization of farmers based on the use of reductant herbicide was analysed in univariate and bivariate, so that they did not represent the characteristics of the categories of farmers simultaneously. This study did not include coffee production and farmers' income variables.

By using a multiple linear regression model, the qualitative variable in the form of the category of reductant herbicide use did not have a significant effect on net income as the dependent variable. This study used data from 136 respondents of Pagaralam coffee farmers with 21 variables, including production and income variables. Variables that have a significant effect on net income are gross income, land maintenance costs, estimated yields, and tree age [25]. Furthermore, multivariate analysis is needed to determine the interdependence among the variables studied, including using cluster analysis and also two groups analysis.

PCA (Principal Component Analysis) is used as an initial analysis, for example, the comparison of the profiles of songket craftsmen in three centers (sub-districts) was carried out using biplot analysis [26] and groups analysis [27]. Groups analysis is an analysis to compare the set of Principal Components (PCs) of PCA results from two or more data matrices (groups), so that the source of variations determines the similarity or dissimilarity of objects between groups [28]. Comparison of subsets of PC sets can represent *p*-dimensional space in the same variables measured for two groups of observations. The PC comparison for two individual groups is an eigenvector comparison of the covariance matrix in the methodology developed by Krzanowski [29]. Next, [30] discusses the hypothesis testing procedure for the subspace spanned by the eigenvectors that form a linear combination of the first to fourth PC represents in the two groups. Subspace comparison which is a form of geometric representation of the group can produce an image [31-32]. [33]compares the variation of personality on 7 freshwater fish species quantitatively by using the PC comparison of the covariance matrix structure.

Cluster analysis can be used to group objects and also variables, which are represented in the form of a dendogram. A group of objects can be characterized by one or more variables. Variables that are closely correlated will form a cluster [34]. The use of cluster analysis on nutritional grouping on the diet menu for diabetics can be seen in [35]. The application of cluster analysis is also used to determine the characteristics of groups of coffee-producing provinces in Indonesia [36] and coffee-producing districts/municipalities groups in South Sumatra [37].

The comparison between the two variable subspaces from the results of the groups analysis was studied further when it is compared to the graphical results of the cluster analysis. The purpose of this study was to compare the characteristics of Pagaralam coffee farmers based on the use of reductant herbicide using two groups analysis and cluster analysis. The characteristics of the farmers were examined on 165 respondents and included 17 variables, namely: farmer identity (i.e. age, length of business in coffee farming, and education), coffee production, and farmers' income, use of workers, and land productivity. The results of the groups analysis were analyzed further by also comparing the variable clusters on the data matrix of all respondents, users and non-users of reductants, so the results show the characteristics of respondents from each category of farmers. The clusters formed are based on their similarity level. The similarity and dissimilarity of the characteristics of the respondent categories can represent the influence of the background of the respondent's identity, land identity, and the impact of land care culture with the use of reductant herbicide on the economic side of farmers.

2. RESEARCH METHODS

Respondents are coffee farmers who own land and run their own coffee farming business in South Dempo District and Dempo Tengah District, Pagaralam Municipality, South Sumatra Province. Respondents were taken by purposive sampling. The sample of respondents is divided into 2 categories, namely users and non-users of reductant herbicide. Reductant users are respondents who used reductant herbicides for more than 1 year with minimal 3 times of use. Meanwhile, non-users are respondents who did not use and include those who had just started using reductants. The steps taken in this research are as follows:

1. Arrange a data matrix of the two categories of respondents.
1.1 Perform descriptive statistics of each variable in both categories of respondents.
1.2 Perform mean difference test and variance ratio test on each variable.
1.3 Compile a data matrix.
2. Determine the correlation matrix of each data matrix of the two categories of respondents.
3. Perform PCA (Principal Component Analysis) on each data matrix.
4. Represent the first two PCs from Step 3 in biplot form.
5. Perform two-group analysis (Krzanowski, 1996):
   5.1 Tabulate the first 3 PCs from Step 3.
   5.2 Define the matrix $L^T_{(p\times k)}=(l_{ij})$ and $M^T_{(p\times k)}=(m_{ij})$, where l_{ij} dan m_{ij} are the coefficients of the first k PC linear combination on the comparison between 2 categories of respondents (or two groups). The value of k is based on the comparison dimension to be analyzed, namely k = 1, 2, 3.
   5.3 Determine the eigenvalues λ_i and their corresponding eigenvectors $\mathbf{a}_i$ from the matrix $N_{(k\times k)} = LM^TML^T$.
   5.4 Determine the size of the angle $\cos^{-1}\sqrt{\lambda_i}$; where λ_i is the ith largest eigenvalue of $N_{(k\times k)}$.

5.5 Determine the bisector $\boldsymbol{c}$ with the equation

$$\boldsymbol{c}_{i\ (p\times 1)} = \{2(1+\sqrt{\lambda_i})\}^{-\frac{1}{2}}\left(I + \frac{1}{\sqrt{\lambda_i}}M^T M\right)\boldsymbol{b}_i;\ i = 1,\ \ldots,\ k \text{ where } \mathbf{b}_i = L^T\ \mathbf{a}_i. \quad (1)$$

5.6 Interpretation of results.

6. Perform cluster analysis with complete, centroid, average and single linkage methods on each data matrices of users, non-users and also a combination of both.
7. Interpret the cluster output from Step 6.
8. Interpretation of overall results.

The steps in this research were using the Minitab 19 software.

3. RESULTS AND DISCUSSION

In this study, respondents were divided into 2 categories, namely users and non-users of reductants. But in reality, the use of reductants can be felt its impact on the land after going through more than 1 harvest period (or it can be said more than 1 year). So, the reductant users can be divided into 2, namely users who have applied reductant herbicides for more than 1 year (with minimum of 3 times applications) and users who have applied for less than 1 year at the time of this research. So, a comparison was made between two categories of respondents, namely users (with notation 1) and a combined category of respondents who have just used and are not users (or they are called as non-users and denoted by 0).

In this research, the number of respondents consisted of 84 non-users and 81 users. The category of respondents is also referred to as a group. There are 17 variables used, so the data matrix of the 2 categories of respondents are 84 × 17 for non-user data and 81×17 for user data. These variables were selected from 28 variables at previously studied, where they had higher PC coefficients than other variables in the initial PCA. The mean and standard deviation of each variable, as well as the results of hypothesis testing with mean difference and variance ratio tests can be seen in Table 1. Descriptive statistics of variables with significantly different mean and variance in the two respondents can be represented as the boxplot in Figure 1.

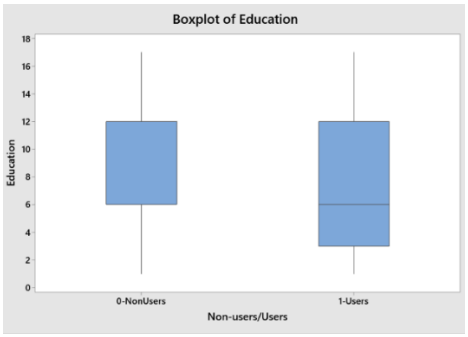
Table 1. Hypothesis testing for difference mean and ratio of variance in two categories of respondents

No.	Variable	Non-users/Users	*N*	Mean	StDev	Z_{count}	F_{count}	Description of Z and F tests
1	Ages	0	84	45,01	11,80	0,09	1,25	Accept H_0
		1	81	44,86	10,57			
2	Education	0	84	10,167	3,957	4,31	1.39	**Reject H_0**
		1	81	7,259	4,658			
3	Length of farming experience	0	84	21,79	12,14	-1,17	1,20	Accept H_0
		1	81	23,91	11,07			
4	Land area	0	84	1,1786	0,5482	-1,60	1.87	Accept H_0
		1	81	1,3426	0,7494			
5	Number of tress	0	84	3662	1700	-0,81	2.18	* Accept H_0
		1	81	3933	2512			
…	…							
15	Length of harvest period	0	84	2,4405	0,5881	-4,31	1,08	**Reject H_0**
		1	81	2,8272	0,5655			
16	Land productivity	0	84	991,8	447,1	1,20	1,27	Accept H_0
		1	81	912,8	397,1			
17	Production average	0	84	3280	1802	-0,10	1,56	Accept H_0
		1	81	3305	1442			

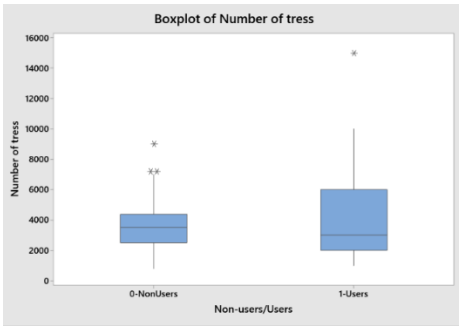
*Note: The critical Z for α/2 = 5% is 1.65; α /2=2.5% is 1.96. The critical F value uses α = 5%. *Meaningly reject H_0 on the F test. The two-tailed hypothesis test on H_0 states that the mean of the two populations is the same. The two populations are assumed to be independent with the Z test statistic.* $Z_{count} = \frac{\overline{x_1}-\overline{x_2}}{\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}}$ *and* $F_{count} = = \frac{s_1^2}{s_2^2}$. *In the value of F_{count}, the larger sample variance is placed in the numerator, while the smaller sample variance is placed in the denominator.*

The mean difference test in Table 1, with = 5%, resulted that only the variables of Education, Age of tree, Frequency of herbicide use, Number of TL, and Length of harvest period have significantly different means between the two categories of respondents. Reductant users had means significantly higher than non-

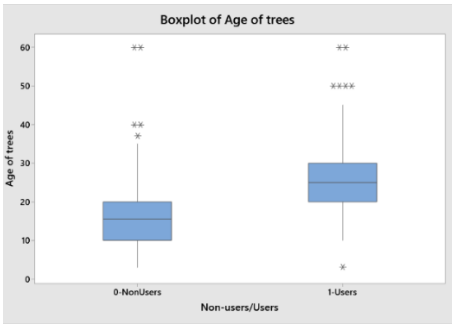
users on these variables, except for the Education variable. While the results of the variance ratio test, reductant users also have a significantly higher variance than non-users in the Number of trees and the Number of TLs variables. For variables whose means are significantly different, they can also occur if the two categories of respondents have almost the same mean and variance of the variables, such as the Frequency of herbicide use and the Length of the harvest period variables.



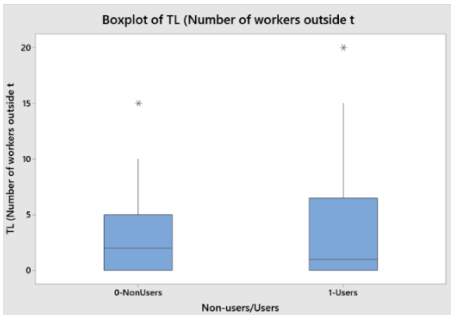
(1) Reject H_0 on mean difference test for Education variable



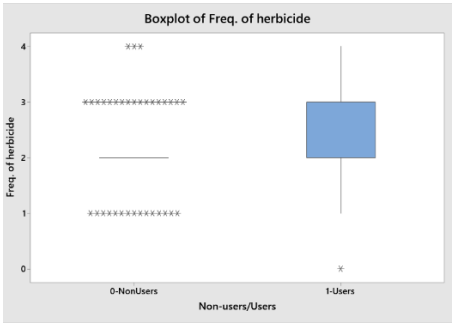
(2) Reject H_0 on variance ratio test for Number of tress variable



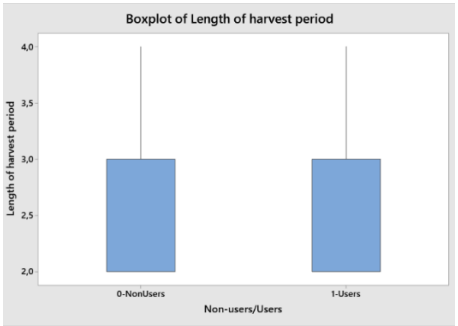
(3) Reject H_0 on mean difference test for Age of trees variable



(4) Reject H_0 on mean difference and variance ratio tests for TL variable



(5) Reject H_0 on mean difference test for Freq. of herbicide variable



(6) Reject H_0 on mean difference test for Length of harvest period variable

Figure 1. Boxplot of the variables that result the reject of H_0 in the mean difference and the variance ratio tests

The boxplot in Figure 1 shows that the median, Q1, Q3, and variance values of the tree age, frequency of herbicide, number of TL, and length of harvest period variables in user data are higher than non-user data. But in Number of TL, the median value in the non-user data is higher than the user data. Based on the correlation matrix of 17 variables in each category of respondents, variables that have a high correlation value (i.e. more than 0.7) can be recapitulated as shown in Table 2.

The correlation between the number of trees and the land area for non-users and users are respectively 0.88 and 0.91. They can be interpreted that the wider land area is in the same as the number of trees, the more trees. This can be related to the farmer's assumption that the more trees, the higher the production. Because this coffee field is partly inherited, there is a culture of adding coffee trees (known as 'sulam') among the existing coffee trees. Farmers also often do not take care their coffee trees optimally, especially those that are old, so that the production of coffee trees is also not optimal. In the non-user group, the number of trees variable is highly correlated with land area, gross income, and net income. Meanwhile, for users, the number of trees is highly correlated with land area, coffee bean production, and total harvest.

Table 2. Variables that have a high correlation value in each group

Variable	Non-users	Users
Length of farming experience	Age (0.89)	Age (0.78)
Number of trees	Land area (0.88)	Land area (0.91)
Coffee bean production		Number of trees (0.73)
	Estimated yield (0.92)	Estimated yield (0.87)
Total harvest		Number of trees (0.75)
	Estimated yield (0.77)	Estimated yield (0.86)
	Coffee bean production (0.80)	Coffee bean production (0.97)
Gross income	Number of trees (0.72)	
	Estimated yield (0.84)	Estimated yield (0.85)
	Coffee bean production (0.88)	Coffee bean production (0.93)
	Total harvest (0.71)	Total harvest (0.87)
Net income	Number of trees (0.71)	
	Estimated yield (0.83)	Estimated yield (0.86)
	Coffee bean production (0.87)	Coffee bean production (0.91)
	Total harvest (0.71)	Total harvest (0.86)
	Gross income (0.98)	Gross income (0.92)
Production average (kg/10^4 trees)	Land productivity (0.86)	

Correlation between variables can also be represented in the form of a biplot as a form of graphical representation of PCA results in 2 dimensions space. Objects in 17 dimensions space are reduced to 2 dimensions space using PCA result from the correlation matrix. The coefficients of PC 1, PC 2, and PC 3 from each group can be seen in Table 2.

Table 2. The coefficients of PC 1, PC 2, and PC 3 in the two groups of respondents

Variable	PC coefficient in the non-user group			PC coefficient in the user group		
	PC 1	**PC 2**	**PC 3**	**PC 1**	**PC 2**	**PC 3**
	*38.3%	16.1%	10%	38.4%	14.6%	11.8%
Age	0.214	-0.234	0.237	0.114	**0.335**	**0.414**
Education	-0.119	0.279	**-0.340**	-0.070	-0.299	**-0.370**
Length of farming experience	0.236	-0.148	0.235	0.081	**0.335**	**0.395**
Land area	0.286	0.283	0.091	0.296	0.252	-0.241
Number of trees	0.290	**0.333**	0.095	**0.323**	0.229	-0.206
Tree age	0.127	-0.286	**0.351**	0.067	0.123	0.282
Estimated yield	**0.348**	0.026	-0.145	**0.342**	-0.147	-0.018
Frequency of herbicide use	0.089	-0.109	0.208	-0.071	0.026	-0.161
Coffee bean production	**0.358**	0.049	-0.141	**0.377**	-0.112	0.007
Total harvest	**0.333**	-0.121	-0.222	**0.378**	-0.101	0.010
Land maintenance costs	0.248	0.017	0.107	0.226	-0.113	-0.064
Gross income	**0.359**	0.117	-0.125	**0.347**	-0.211	0.025
Net income	**0.346**	0.130	-0.171	**0.340**	-0.153	0.060
Number of workers outside the family (TL)	0.096	-0.086	0.255	0.242	0.008	-0.117
Length of harvest period	0.031	-0.021	**0.445**	0.153	0.147	0.205
Land productivity (kg/10^4 m^2)	0.095	**-0.500**	-0.297	0.025	**-0.467**	**0.347**
Production average (kg/1000 trees)	0.069	**-0.504**	**-0.302**	-0.063	**-0.435**	**0.380**

Note: Numbers in bold indicate the higher coefficient value (dominant) in each PC.
*The * sign represents the contribution of variation represented by the PC.*

Every object in each group (user and non-user data matrices) initially resides in a 17 dimensions space. After the groups analysis is performed, they are represented in 1 dimension, 2 dimensions, and 3 dimensions spaces that is depending on the number of first PCs used. The bisector is "a mean vector" between two PCs of both of subgroups (or both subspaces of variables), so the angle obtained is the angle formed by the bisector with each PC. The angles in the first bisector in all comparison of subspace dimensions are 20.53^0, 16.9^0, and 15.4^0, so the coefficient of variables on this bisector determines the similarity or dissimilarity of the two groups. The dominant variables determining the similarity between the two groups are Land area, Number of trees, Coffee bean production, estimated yield, Total harvest, Gross income, and Net income. In the 2 dimensions space, the second bisector angle formed is 72.6^0 and in the 3 dimensions space it is 24.4^0, so that land productivity and Production average are variables that show dissimilarity between the two groups. While

in the 3 dimensions space, the angle formed from the third bisector is 40^0, so that the variables of respondent's Age, Education, Length of farming experience, Age of trees, and Length of harvest also determine the dissimilarity of the two groups.

The variance ratio test for the Number of trees variable results different variances in the data of the two categories of respondents. This can be related to the high correlation value between the number of trees and the other variables in each category of respondents. For non-users, it is found that the Number of trees is highly correlated with Land area, Gross income, and Net income. Meanwhile, for users, the Number of trees is highly correlated with Land area, Total harvest, and Coffee bean production.

The correlation between variables in the biplot results shows that for non-users, land maintenance is related to coffee bean production and Estimated yield. In this case, the high land maintenance cost on non-user farmers is in line with the increase in coffee production. Meanwhile, for users, production (consisting of total harvest, coffee bean production, and estimated yield) has a high positive correlation with net income. In this case, high production on users is in line with high net income as well.

If the results of the biplot are related to the results of the groups analysis, then the dominant variable that characterizes each biplot is also the dominant variable determining the similarity of the two categories of respondents, namely: Coffee bean production, estimated yield, and total harvest. If the results of the mean difference test are associated with the results of group analysis, the following are obtained:

(i) There are two variables whose means of two categories of respondents are significantly different. But they are not dominant in determining the dissimilarity of the two categories of respondents, i.e. Frequency of herbicide use and Number of TL. This is in line with the PCA results, where these variables are not the dominant variables generate each subspace.

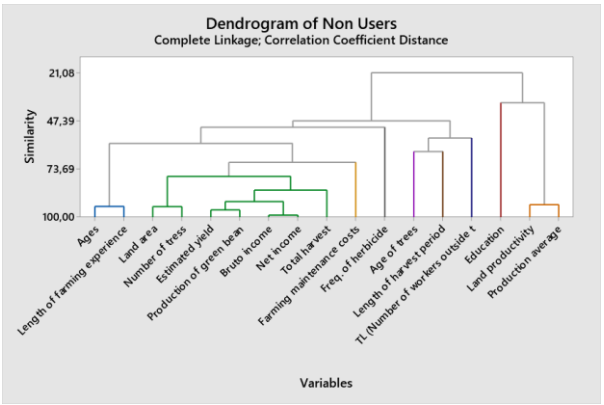
(ii) There are 3 variables whose means of two categories of respondents are significantly different. They are dominantly in determining the dissimilarity of the two categories of respondents, i.e. Education, Age of trees, and Length of harvest period.

(iii) there are 4 variables whose means of two categories of respondents are not significantly different. They are dominantly in determining the dissimilarity of the two categories of respondents, i.e. respondent's age, length of farming experience, land productivity, and Production average. These four variables are also in line with the results of PCA, where on PC 3, these variables dominantly characterize the subspace of the users' group.

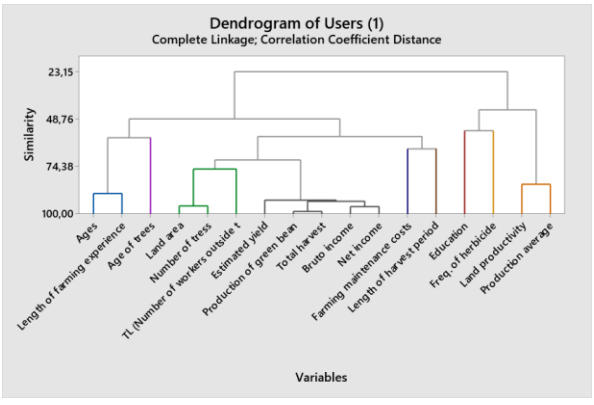
(iv) There are 7 variables whose means of two categories of respondents are not significantly different. They are dominantly in determining the similarity of the two categories of respondents, i.e. land area, number of trees, Production of green beans, Yields estimation, Total harvest, Gross income, and Net income. This is also in line with the correlation value between these variables. Although the variance of number of trees is significantly different in the two groups.

(v) Farming maintenance costs variable tends not to be dominant in determining the similarity or dissimilarity of the two groups of respondents.

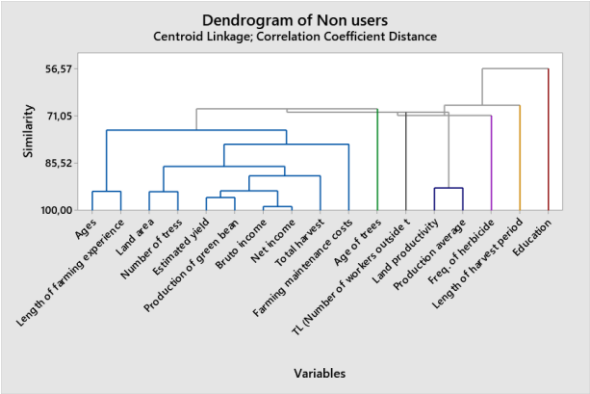
Furthermore, in each group, cluster analysis was performed at 75% similarity using complete linkage, single linkage, average linkage, and centroid linkage methods. The dendograms for each category of respondents and the combine all respondents shown on Figure 3 are only the results of complete linkage and centroid linkage. While the number of clusters and variables that characterize each cluster can be seen in Table 4.



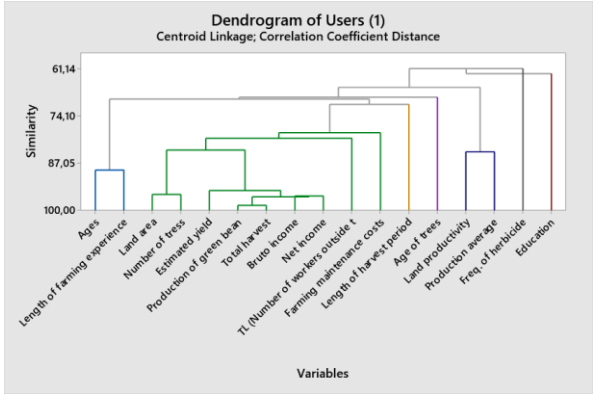
a. The output of the complete linkage method on the data of non-users



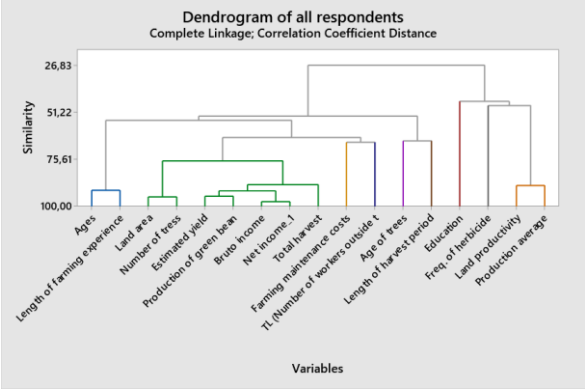
b. The output of the complete linkage method on the data of users



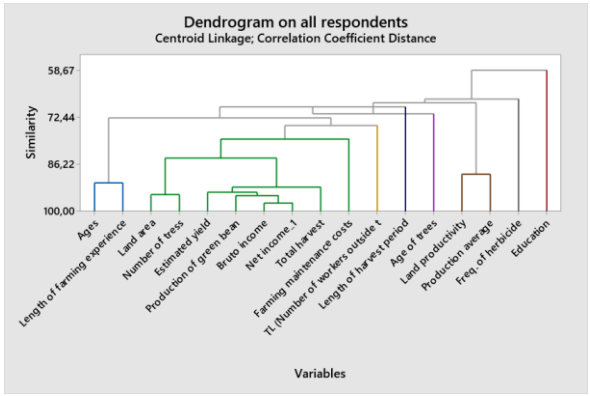
c. The output of the centroid linkage method on the data of non-users



d. The output of the centroid linkage method on the data of users



e. The output of the complete linkage method on the data of all respondents



f. The output of the centroid linkage method on the data of all respondents

Figure 3. Dendogram on the data of users, non-users, and all respondents

In all the data of users, non-users and combined all respondents, the number of clusters formed from the results of complete linkage is more, namely there are 9 clusters. The four methods result the same 5 clusters, with 1 of them consisting of two variables, namely on the plantation productivity cluster. This cluster consists of the Land productivity and Average production variables. Each method results at least 4 clusters with only 1 variable member in a cluster. Those clusters are Education (X_2), Age of trees (X_6), Frequency of herbicide use (X_8), and Length of harvest period (X_{15}). Meanwhile, the Farming maintenance costs (X_{11}) and Number of TL (X_{14}) variables in the results of several methods can join in the other clusters.

In the non-user data, the number of clusters with the same members from the outputs of the four methods is 6 clusters. One of which is characterized by 2 variables (i.e. Land productivity and Production average), namely the cluster of plantation productivity. While 1 cluster on the output of centroid linkage (that is in cluster 1), its members become 3 clusters on the output of each complete and average linkage, namely the cluster of land identity, the cluster of production and income, and the cluster of Farming maintenance cost. In the output of the single linkage method, Farming maintenance costs (X_{11}) join the cluster of production and income. In each method, the Number of TL (X_{14}) forms its own cluster.

In user data, there are 6 clusters whose members are the same as the outputs of the four methods. Two of which are characterized by 2 variables, namely the cluster of respondent identity (that consists of Age and Length of farming experience variables) and the cluster of plantation productivity (that consists of Land productivity and Production average variables). While 1 cluster (that is in cluster 3) on the outputs of centroid and single linkages, its members become 3 clusters on the output of complete linkage. They are the cluster of land identity and TL (that consists of land area, number of trees, and TL variables), the cluster of production and income, and the cluster of farming maintenance costs. The cluster 3 in the output of centroid linkage forms 2 clusters in the output of average linkage. The most cluster member variables are in the cluster of production factors resulting from the output of centroid linkage, which includes farmer and land identities, production yields, income, and maintenance costs including the use of TL. So, in user data, Number of TL (X_{14}) joins cluster 3, while Farming maintenance costs (X_{11}) can form its own cluster on the output of complete and average linkages.

Table 4. Recapitulation of variables on each cluster in the output of complete, single, average, and centroid linkage methods

Cluster	Data of non-user data			Data of users			Data of all respondents	
	Complete Average	Centroid	Single	Complete	Centroid Single	Average	Complete Average	Centroid Single
Amount * > 1 variable	9 3	7 2	8 3	9 4	7 3	8 3	9 3	8 3
Cluster 1	X_1 X_3	X_1 X_3 X_4 X_5 X_7 X_9 X_{10} $\boldsymbol{X_{11}}$ X_{12} X_{13}	X_1 X_3	X_1 X_3	X_1 X_3	X_1 X_3	X_1 X_3	X_1 X_3
Cluster 2	X_2	X_2	X_2	X_2	X_2	X_2	X_2	X_2
Cluster 3	X_4 X_5 X_7 X_9 X_{10} X_{12} X_{13}		X_4 X_5 X_7 X_9 X_{10} $\boldsymbol{X_{11}}$ X_{12} X_{13}	X_4 X_5 $\boldsymbol{X_{14}}$	X_4 X_5 X_7 X_9 X_{10} $\boldsymbol{X_{11}}$ X_{12} X_{13} $\boldsymbol{X_{14}}$	X_4 X_5 X_7 X_9 X_{10} X_{12} X_{13} $\boldsymbol{X_{14}}$	X_4 X_5 X_7 X_9 X_{10} X_{12} X_{13}	X_4 X_5 X_7 X_9 X_{10} $\boldsymbol{X_{11}}$ X_{12} X_{13}
Cluster 4	X_{14}	X_{14}	X_{14}	X_7 X_9 X_{10} X_{12} X_{13}			X_{14}	X_{14}
Cluster 5	$\boldsymbol{X_{11}}$			$\boldsymbol{X_{11}}$		$\boldsymbol{X_{11}}$	$\boldsymbol{X_{11}}$	
Cluster 6	X_6	X_6	X_6	X_6	X_6	X_6	X_6	X_6
Cluster 7	X_8	X_8	X_8	X_8	X_8	X_8	X_8	X_8
Cluster 8	X_{15}	X_{15}	X_{15}	X_{15}	X_{15}	X_{15}	X_{15}	X_{15}
Cluster 9	X_{16} X_{17}	X_{16} X_{17}	X_{16} X_{17}	X_{16} X_{17}	X_{16} X_{17}	X_{16} X_{17}	X_{16} X_{17}	X_{16} X_{17}

*Note: *Number of clusters that their member > 1 variable*
X_1 : Ages, X_2 : Education, X_3 : Length of coffee frming, X_4 : Land area, X_5 : Number of trees, X_6 : Age of trees, X_7 : Estimated yield, X_8 : Freq. of herbicide use, X_9 : Coffee bean production, X_{10} : Total harvest, X_{11} : Farming maintenance costs, X_{12} : Gross income, X_{13} : Net income, X_{14} : Number of TL, X_{15} : Length of harvest period, X_{16} : Land productivity, X_{17} : Production average

Meanwhile, on the combined data, the outputs of complete and average linkage methods are the same as the output of them on non-user data. On the other hand, the outputs of centroid and single linkage methods are the same as the output of the single linkage method in non-user data. If the output clusters of the four methods in the two categories of respondents are compared, then there are differences in cluster characteristics, namely:

- For non-user data, the cluster of respondent identity join together in cluster of production and income on the output of centroid linkage.
- For non-user data, Number of TL (X_{14}) forms its own cluster. On the other hand, in user data, this variable join into the cluster of land identity or the cluster of production and income.

Other similarities between the two categories of respondent data are:

- Farming maintenance costs on the output of complete and average linkages are separate from clusters related to land identity, production, and income. Meanwhile, for the outputs of centroid and single linkages, the maintenance costs variable is included in the cluster.
- Variables related to land identity, production and income, can form the same cluster.

- There are 5 clusters with the same members from the outputs of four methods on the three data, namely the clusters of the Education, the Tree Age, the Freq. of Herbicide use, the length of harvest period, and the plantation productivity.

Based on the results obtained from the analysis steps, variables from the two categories of respondents that have significantly different mean and variance values do not necessarily determine the dissimilarity or similarity of the two categories of respondents, and vice versa. If the results of the mean difference test are compared with the results of the groups analysis, then only the frequency of herbicide use and the number of TL are non-dominant variables which determine the similarity of the two groups. This is also in line with the PCA results regarding dominant or non-dominant variables that characterize subspaces. The variables that dominantly determine the similarity between the two groups are in line with the results of the correlation between the variables on the biplot results of each group. The variables that dominantly characterize each subspace of group on the biplot are also the dominant variables that determine the similarity of the two groups, namely Coffee bean production, Estimated yield, and Total harvest.

The variables on the result of groups analysis that dominantly characterize the similarity of the two categories of respondents are the same as the variables whose mean difference test on two categories of respondent are not significantly different. Those variables consist of 7 variables, namely land area, number of trees, Coffee bean production, estimated yield, Total harvest, Gross income, and Net income. But the variance of number of trees is significantly different in both groups.

Meanwhile, the variables that dominantly characterize the dissimilarity of the two categories of respondents, which also have significantly different on the result of mean difference test, namely there are 3 variables consisting Education, Tree Age, and Length of Harvest. Other variables that also determine the dissimilarity of the two groups, but on the results of the mean difference test they are not significantly different, are Age of respondents, length of farming experience, land productivity, and Production average. While the Farming maintenance costs variable is not dominant in determining the similarity or dissimilarity of the two groups.

Overall, the results of the groups analysis are also in line with the cluster analysis. The variables that determine the similarity of the characters of the two categories of respondents are variables that are incorporated in one cluster. Meanwhile, the variables that determine the dissimilarity of these characters are variables that characterize a separate cluster that is separated from other clusters. Cluster separation is caused by the low level of similarity.

4. CONCLUSIONS

According to the results of the two groups analysis and cluster analysis, the characters dominantly tend to characterize the similarity in comparison between the categories of non-users and users are the variables that join together into one cluster, namely: land area, number of trees, coffee bean production, estimated yield, total harvest, Gross income, and Net income. While the characters that dominantly characterize the dissimilarity between the two categories of reductant use are the variables that form 5 separate clusters. The five clusters are the plantation productivity cluster (that combined land productivity and average production variables), the respondent identity cluster (combined age of farmers and length of farming experience variables), Education, Age of the tree, and Length of harvest. In each group of data matrix, the four methods of cluster analysis resulted in 5 clusters. But, the results of centroid linkage in the non-user data matrix was obtained that the respondent's identity cluster was joined by 1 cluster of land identity, production, and income. In further research, the comparison of characteristics between the two groups of respondents can also be done by a classification process. This classification is done by allocating respondents to one of the groups, for example by discriminant analysis.

Several variables that affect the differences in the characteristics of the two categories of respondents relate to the steps that need to be considered in land care, so that coffee plants remain healthy and produce optimally and without pests and diseases. The production of a lot of cherries does not necessarily produce comparable coffee beans production. This is in accordance with the fact on the field that older coffee plants, which have a higher density, should require more intensive land and crop care.

ACKNOWLEDGEMENT

The research of this article was funded by DIPA of Public Service Agency of Universitas Sriwijaya 2022, SP DIPA-023.17.2.677515/2022, on December 13, 2021. In accordance with the Rector's Decree Number: 0109/UN9.3.1/SK /2022, on April 28, 2022. The authors would also like to thank PT. Pandawa Agri Indonesia (PAI), especially the South Sumatra team: Mr. Ahmad Surkati, Rici Wijaya, Kania, and Nando, who have assisted in obtaining data in the field.

REFERENCES

[1] D. J. P. Ditjenbun, *Statistik Perkebunan Unggulan Nasional 2019-2021*. Direktorat Jenderal Perkebunan Kementerian Pertanian, 2020.
[2] Y. M. P. Putra, "Kopi Robusta Pagaralam Raih Sertifikat Indikasi Geografis," *Republika.co.id*, 2020. https://www.republika.co.id/berita/qiia77284/kopi-robusta-pagaralam-raih-sertifikat-indikasi-geografis (accessed Jun. 06, 2022).
[3] Asnadi, "Pemerintah Terbitkan Sertifikat Indikasi Geografis untuk Kopi Robusta Pagaralam," *TimesIndonesia*, 2020. https://www.timesindonesia.co.id/read/news/301004/pemerintah-terbitkan-sertifikat-indikasi-geografis-untuk-kopi-robusta-pagaralam (accessed Jun. 06, 2022).
[4] BPS, *Kota Pagar Alam dalam Angka 2021*. BPS Kota Pagaralam, 2021.
[5] Irmeilyana, Ngudiantoro, and D. Rodiah, "Deskripsi Profil dan Karakter Usaha Tani Kopi Pagar Alam Berdasarkan Descriptive Statistics dan Korelasi," *Infomedia*, vol. 4, no. 2, pp. 60–68, 2019, doi: http://dx.doi.org/10.30811/jim.v4i2.1534.
[6] Irmeilyana, Ngudiantoro, and D. Rodiah, "Application of Simple Correspondence Analysis to Analyze Factors that Influence Land Productivity of Pagar Alam Coffee Farming," 2019.
[7] Ngudiantoro, Irmeilyana, and M. N. Samsuri, "Binary Logistic Regression Modeling on Net Income of Pagar Alam Coffee Farmers," *Int. J. Appl. Sci. Smart Technol.*, vol. 2, no. 2, pp. 47–66, 2020.
[8] Irmeilyana, Ngudiantoro, M. N. Samsuri, and B. Suprihatin, "Logistic regression model on land productivity of Pagar Alam coffee farming," in *J. Phys.: Conf. Ser. 1943*, 2021, pp. 1–12, doi: doi:10.1088/1742-6596/1943/1/012135.
[9] Irmeilyana, Ngudiantoro, and D. Rodiah, "Correspondence analysis pada hubungan faktor-faktor yang mempengaruhi pendapatan petani kopi Pagaralam," *BAREKENG J. Ilmu Mat. dan Terap.*, vol. 15, no. 1, pp. 179–192, 2021, doi: https://doi.org/10.30598/barekengvol15iss1year2021.
[10] J. Kath, V. Mittahalli, Byrareddy, S. Mushtaq, A. Craparo, and P. Mario, "Temperature and rainfall impacts on robusta coffee bean characteristics," *Clim. Risk Manag.*, vol. 32, no. 100281, pp. 1–15, 2021, [Online]. Available: www.elsevier.com/locate/crm.
[11] G. Schroth, L. derach Derach, D. S. B. Cuero, J. Neilson, and C. Bunn, "Winner or loser of climate change? A modeling study of current and future climatic suitability of Arabica coffee in Indonesia," *Electron. Suppl. Mater.*, vol. 15, pp. 1473–1482, 2015, doi: :10.1007/s10113-014-0713-x.
[12] I. Nzeyimana, A. E. Hartemink, and J. De Graaff, "Coffee farming and soil management in Rwanda," *Outlook Agric.*, no. March 2019, pp. 47–52, 2013, doi: 10.5367/oa.2013.0118.
[13] TechnoServe, *Improving the Productivity and Sustainability of Smallholder Coffee Farmers in Guatemala (A Case Study of TechnoServe's coffee project in Sololá, Chimaltenango, and Socatepéquez 2012 – 2017)*. Washington, 2018.
[14] IDH," Sustainable Coffee Program: A business case for sustainable coffee production," no. March, 2014, [Online]. Available: http://exchange.growasia.org/indonesia-business-case-sustainable-coffee-production.
[15] Lizawati, E. Kartika, and Gusniwati, "PPM Kelompok Tani Kecamatan Betara : Aplikasi Teknologi Pupuk Organik Mikotriderm Dalam Penerapan Good Agricultural Practices Pada Budidaya Kopi Liberika Tungkal Jambi," *J. Karya Abdi Masy.*, vol. 3, no. 1, pp. 22–28, 2019, [Online]. Available: http://download.garuda.ristekdikti.go.id/article.php?article=1053114&val=14365&title=
[16] Dinas Pertanian Kabupaten Buleleng, "Pengertian dan Konsep Pertanian Berkelanjutan," *Dinas Pertanian*, 2019. https://distan.bulelengkab.go.id/informasi/detail/artikel/pengertian-dan-konsep-pertanian-berkelanjutan-22 (accessed Jun. 06, 2022).
[17] D. Febriana, E. G. Ayu, H. Sya'bani, and S. S. D, "Pengelolaan Pertanian Berkelanjutan," *Klinik Agromina Bahari FP UGM*, 2021. https://kab.faperta.ugm.ac.id/2021/05/25/pengelolaan-pertanian-berkelanjutan/ (accessed Jun. 06, 2022).
[18] N. R. Endah, M. Masahiko, and Y. Purbokusumo, "DEVELOPMENT OF SUSTAINABLE AGRICULTURE IN INDONESIA Case Study: Economic and Environmental Benefit of Organic Farming in Java Island," UGM, 2017.
[19] Triharso, "Kerugian Akibat Gulma di Perkebunan Kopi dan Cara Mengatasinya," 1994. https://www.cr1coffee.com/sub/kerugian-akibat-gulma-di-perkebunan-kopi-dan-cara-mengatasinya/ (accessed Jun. 25, 2021).
[20] Irmeilyana, Ngudiantoro, and S. I. Maiyanti, "Socialization of sustainable Pagar Alam Coffee Farming using herbicide reductors," *Abdimas*, vol. 7, no. 2, pp. 309–318, 2022, [Online]. Available: https://jurnal.unmer.ac.id/index.php/jpkm/article/view/6372.
[21] KRJOGJA.com, "Weed Solut-ion' Dukung Pertanian Berkelanjutan dan Ramah Lingkungan," 2021. https://www.krjogja.com/angkringan/gaya-hidup/teknologi/weed-solut-ion-dukung-pertanian-berkelanjutan-dan-ramah-lingkungan/ (accessed Mar. 07, 2021).
[22] T. Surya, "Penurunan dosis penggunaan herbisida bisa hingga 50%," *AGRINA-ONLINE.COM*, 2021. http://www.agrina-online.com/detail-berita/2021/02/18/56/7314/biaya-perkebunan-hemat-hingga-40-dengan-reduktan-herbisida (accessed Mar. 07, 2021).
[23] K. Sarirahayu and A. Aprianingsih, "Strategy to Improving Smallholder Coffee Farmers Productivity Literature Study," vol. 11, no. 1, pp. 1–9, 2018.
[24] I. Irmeilyana, Ngudiantoro, and S. I. Maiyanti, "Hypothesis testing in the study of the characteristics of Pagaralam coffee farmers as herbicide reductant users," 2021.

[25] Irmeilyana, Ngudiantoro, and S. I. Maiyanti, "Regression Model on Pagaralam Coffee Farmers' Income with the Influence of The Use of Herbicide Reductant Variable," *BAREKENG J. Ilmu Mat. dan Terap.*, vol. 16, no. 2, pp. 409–420, 2022, doi: https://doi.org/10.30598/barekengvol16iss2pp409-420.
[26] I. Irmeilyana and A. Desiani, "Analisis Perbandingan Profil Pengrajin di Tiga Sentra Kerajinan Tenun Songket Palembang," *J. Infomedia*, vol. 3, no. 2, pp. 58–63, 2018, doi: 10.30811/jim.v3i2.714.
[27] Irmeilyana, Ngudiantoro, A. Desiani, and D. Rodiah, "Aplikasi Groups Analysis pada Perbandingan Tiga Sentra Kerajinan Tenun Songket Palembang," *Infomedia*, vol. 4, no. 1, pp. 8–14, 2019, doi: 10.30811/jim.
[28] W. J. Krzanowski, *Principle of Multivariate Analysis, A User's Perspective*. New York: Oxford University Press, 1990.
[29] I. T. Jolliffe, *Principal Component Analysis*, 1st ed. New York: Springer Verlag, 1986.
[30] T. Fujioka, "An approximate test for common principal component subspaces in two groups," *Ann. Inst. Stat. Math*, vol. 45, no. 1, pp. 147–158, 1993.
[31] S. N. Afriat, "Orthogonal and oblique projectors and the characteristics of pairs of vector spaces," in *Mathematical Proceedings of the Cambridge Philosophical Society*, 2008, pp. 800–816.
[32] K. Fukui and A. Maki, "Difference Subspace and Its Generalization for Subspace-Based Methods," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 11, pp. 2164–2177, 2015.
[33] S. J. White, D. J. Pascall, and A. J. Wilson, "Towards a comparative approach to the structure of animal personality variation," *Behav. Ecol.*, vol. 31, no. 2, pp. 340–351, 2020.
[34] Irmeilyana, R. Sania, and K. Putahera, "Analisis Karakteristik Mahasiswa Berdasarkan Nilai Kelompok Mata Kuliah dengan Menggunakan Analisis Cluster K-Means," vol. 4, no. 1, pp. 978–979, 2018.
[35] Irmeilyana, Indrawati, and M. D. Putri, "Karakteristik Menu Makanan pada Penderita Diabetes Melitus Berdasarkan Kandungan Gizi dan Harga Makanan Menggunakan Analisis Biplot dan Analisis Klaster," vol. 4, no. 1, pp. 978–979, 2018.
[36] Irmeilyana, Ngudiantoro, A. Desiani, and D. Rodiah, "Deskripsi hubungan luas areal dan produksi perkebunan kopi di Indonesia menggunakan analisis bivariat dan analisis klaster," *Infomedia*, vol. 4, no. 1, pp. 21–27, 2019, doi: http://dx.doi.org/10.30811/jim.v4i2.1534.
[37] Irmeilyana, Ngudiantoro, A. Desiani, and D. Rodiah, "Deskripsi hubungan luas areal dan produksi perkebunan kopi di provinsi sumatra selatan," in *Proseding SEMIRATA BKS PTN Indonesia Barat*, 2019, pp. 74–86, [Online]. Available: http://semiratathe2ndicst.fmipa.unib.ac.id/wp-content/uploads/2019/11/Bidang-Matematika-lengkap.pdf.