Advanced Mathematical Models & Applications
Vol.8, No.3, 2023, pp.565-589

CLUSTER ANALYSIS WITH VARIOUS COMBINATION OF DISTANCE AND LINKAGE FOR MODELING DUMMY VARIABLE PATH ANALYSIS

Solimun[1*], **Ni Wayan Surya Wardhani**[1], **Adji Achmad Rinaldo Fernandes**[1], **Dany Primanita Kartikasari**[2], **Benny Hutahayan**[3]

[1]Department of Statistics, Faculty of Mathematics and Natural Science,
Brawijaya University, Malang, Indonesia
[2]Department of Informatics Engineering, Faculty of Computer Science,
Brawijaya University, Malang, Indonesia
[3]Department of Administrative Science, Faculty of Administrative Science,
Brawijaya University, Malang, Indonesia

Abstract. In this paper, we integrate cluster analysis with path analysis to model compliance behavior for paying the House Ownership Loan at Bank X. A sample size of 200 customers was obtained using simple random sampling, and three distance measures and clustering methods were used. The results suggest that the Mahalanobis-Complete Linkage combination is the most suitable for the path model, and that attitude has a stronger influence on willingness to pay in low clusters, while it has a stronger effect on behavioral intention and compliance behavior in high clusters. Path analysis needs to be integrated with cluster analysis in data with large sample sizes, and this study can be seen as the development of a multigroup moderating variable in path analysis.

Keywords: cluster, path, integration, moderation multigroup, compliance behavior to pay, bank.
AMS Subject Classification: 62P20.
Corresponding author: Solimun, Department of Statistics, Faculty of Mathematics and Natural Science, Brawijaya University, Malang, Indonesia, Tel.: +62812-5275-079, e-mail: *solimun@ub.ac.id*
Received: 3 July 2023; Revised: 24 July 2023; Accepted: 12 August 2023;
Published: 20 December 2023.

1 Introduction

The integration of cluster analysis with path analysis is carried out by statistical modeling on a relatively large sample size (Khalifa et al., 2021; Ersoy et al., 2021). On large samples, maybe there are several parts of the data that show the behavior of the relationship between different variables. Some parts of the data can be obtained through cluster analysis, which is used for grouping various objects into the homogen cluster. Furthermore, the behavior of the relationship between the different variables is modeled through path analysis. This integration is done by path analysis using a dummy variable, which is obtained from the results of cluster analysis. To integrate cluster analysis with path analysis, dummy variables are used in path modeling. This is a development of dummy variable regression analysis, which involves categorical predictor variables (Gujarati, 2003; Afifi & Clark, 1990). The grouping results from cluster analysis are

How to cite (APA): Solimun, Wardhani, N.W.S., Fernandes, A.A.R., Kartikasari, D.P. & Hutahayan, B. (2023). Cluster analysis with various combination of distance and linkage for modeling dummy variable path analysis. *Advanced Mathematical Models & Applications*, *8*(3), 565-589.

used as input into path analysis, with the use of dummy variables. For instance, if 3 clusters are produced in cluster analysis, path analysis modeling would involve 2 dummy variables.

Path analysis is a multivariate statistical method of relatively complex relationships between variables (Dillon & Goldstein, 1984). Path analysis is the development from the regression analysis used to determine the direct and indirect effect of the relationship between several variables. Path analysis is applied quite extensively in research in various fields, including research in the social humanities field. Research in the social humanities field often involves multi variables with relatively large numbers and complex relationships. The statistical method that is often applied to that research is the multivariate statistical method, which is path analysis. The use of such multivariate analysis has a consequence that is needed for the number of samples relatively large, this large sample numbered more than 100 (Hair et al., 2010). As in research on compliance behavior to pay House Ownership Loan at Bank X as follows.

The Theory of Planned Behavior (Ajzen, 1991) is the foundation of a belief perspective that is able to influence someone to carry out specific behavior. This theory explains that attitude towards behavior is an important factor that can predict an action. Based on this theory, compliance behavior to pay is influenced by attitudes, behavioral intentions, and willingness to pay. The modeling in this study is based on this theory. How customer payment compliance behavior can reflect customer preferences, is important to know so that companies can survive in the midst of competition. (Sumardi & Fernandes, 2020). Militancy and a culture of innovation are needed so that banks become more competitive in the future (Wijayanto et al., 2021). In addition, it is necessary to increase the focus on customers in order to improve the company's experience so that it can improve the company's performance and sustainability (Rumahorbo et al., 2022; Siregar et al., 2022).

In data with large samples, it is possible that there are several relationship behaviors between variables. Cluster analysis can be used to map this, by grouping objects based on several variables together (Hair et al., 2010; Johnson & Wichern, 2002; Sharma, 1996). Compliance behavior to pay House Ownership Loans at Bank X has variations in several customer groups. These behavioral variations can be grouped according to their similar behavior using cluster analysis. Each group has different behavioral tendencies, so analysis is needed to accommodate the behavior of each group. This study uses cluster integration modeling with path analysis to accommodate differences in the behavior of each group.

Cluster analysis is a multivariate analysis technique that has the main goal of grouping objects based on their similar characteristics. The purpose of cluster analysis is to group a number of objects into a cluster so that each cluster will load as much may the data have characteristics homogeneous, as well as between clusters have heterogeneous characteristics (Hair et al., 2010). The basic concept of measurement in cluster analysis is distance measurement and similarity measurement (Dicky, 2016). There are several methods of measuring distances that are often used, including Manhattan, Euclidean, Mahalanobis, and Minkowski. There are several clustering methods, including Single Linkage, Ward Linkage, Complete Linkage, and Average Linkage (Hair et al., 2010).

So, it is necessary to integrate cluster analysis and path modeling on the compliance behavior of customers of House Ownership Loan at Bank X at Bank X. By using this modeling, it is expected to accommodate differences in the behavior of the relationship between variables in several groups of House Ownership Loan's customers at Bank X. A combination of three distance measures, namely Manhattan, Euclidean, and Mahalanobis distances, with three clustering methods, namely Single Linkage Method, Complete Linkage method, and Average Linkage method in cluster analysis is used in this study. The results of the cluster analysis are used as a dummy variable in path modeling.

Therefore, this study aims to integrate cluster analysis with path analysis. Integration is carried out through the process of obtaining the best combination of distance and linkage measurements in the clustering of compliance behavior in paying the House Ownership Loan of Bank X customers, where the results are used as a dummy variable in path modeling. This study can be seen as the development of multigroup moderating variables in path analysis.

2 Literature Reviews

2.1 Cluster Analysis

Cluster analysis is a multivariate analysis used to group objects into several clusters (groups) based on the similarity of the observed variables. so those objects that have the same resemblance will be in one cluster and can be compared between objects from different clusters (Marsili & Bödefeld, 2021; Fernandes & Taba, 2019). Cluster Analysis aims to group objects based on their similar characteristics of these objects. Characteristic features of a good cluster are homogeneity (within-cluster), that is high similarity among members in one cluster and heterogeneity (between-cluster), that is high differences between one cluster and another (Wierzchoń & Ktopetek, 2018).

The process of clustering can be done using two methods, namely hierarchical and non-hierarchical (Alharbi & Khalifa, 2021; El-Sanowsy & Atef, 2022; Agama & Repalle, 2022). The Hierarchical Cluster method begins by grouping two objects that have the highest similarity. Then the operation is passed to another object that has second closeness. Thereby so on, so that cluster will form a sort of 'tree' and then produce a clear hierarchy (levels) between objects, from the most similar to the least look alike (Jain & Dubes, 1988; Hair et al., 2010). According to Johnson & Wichern (2002) , there are two approaches in the formation method group on the hierarchical method, namely agglomerative hierarchical methods (hierarchical methods agglomerative) and divisive hierarchical methods (hierarchical methods divisive). The frequent method used is the hierarchical method agglomerative (Kozae et al., 2021; Lone et al., 2021; Lashin & Malibari, 2022).

2.1.1 Linkage Method

There are several grouping methods commonly used in agglomerative hierarchical cluster analysis algorithms, including the following (Fernandes et al, 2020):

a. Single Linkage Method
In the complete linkage method, the distance between clusters is determined by the farthest distance between two objects in different clusters (Johnson & Wichern, 1992). The selection of the farthest distance in the complete linkage method is presented in equation (1).

$$d_{(ij)k} = \min(d_{ik}, d_{jk}) \tag{1}$$

Description:
$d_{(ij)k}$: distances between sub-samples/objects ij and cluster k
d_{ik} : distance of sub-sample/object i and cluster k
d_{jk} : distance of sub-sample/object j and cluster k

b. Complate Linkage Method
In the complete linkage method, the distance between clusters is determined by the farthest distance between two objects in different clusters (Johnson & Wichern, 1992). The selection of the farthest distance in the complete linkage method is presented in equation (2).

$$d_{(ij)k} = \max(d_{ik}, d_{jk}) \tag{2}$$

Description:
$d_{(ij)k}$: distances between sub-samples/objects ij and cluster k
d_{ik} : distance of sub-sample/object i and cluster k
d_{jk} : distance of sub-sample/object j and cluster k

c. Average Linkage Method
   In the average linkage method, the distance between two clusters is considered as the average distance between all members in one cluster and all members of other clusters. The distance formula can be written in equation (3).

$$d_{(ij)k} = \frac{\sum\sum d_{ij}}{N_{ij}N_k} \tag{3}$$

   Description:
   $d_{(ij)k}$: distances between sub-samples/objects ij and cluster k
   d_{ik} : distance of sub-sample/object i and cluster k
   d_{jk} : distance of sub-sample/object j and cluster k

2.1.2 Distance Formulation Method

The distance measurement method is a method that is carried out by measuring the distance between research objects. The smaller the distance between objects, the more homogeneous the object is, and vice versa. There are several methods of measuring distances, including Euclidean distance, City- Block (Manhattan) distance, and Mahalanobis distance (Hair et al., 2010).

a. Euclidean Distance
   Euclidean distance is the most commonly used measurement method. Usually, it can also be referred to as the straight line distance. Euclidean distance between two object points can be thought of as the length of the hypotenuse of a right triangle. Euclidean method distance measures the sum of the squares of the difference in values between two objects using equation (4).

$$d(x_i, x_j) = \sqrt{\sum_{k=1}^{p} (x_{ik} - x_{jk})^2} \tag{4}$$

   Description:
   $d(x_i, x_j)$: the distance between object i and object j
   x_{ik} : data from object i on the k variable
   x_{jk} : data from object j on the k variable
   i, j : $1, 2, ..., n$
   k : $1, 2, ..., p$

b. Manhattan Distance
   Proximity measure using Manhattan distance is the distance between two objects which is the sum of the absolute differences. Formulation of this distance is the easiest but can produce clusters that are invalid if the variables used have a high correlation. The distance formula can be written like equation (5).

$$d(x_i, x_j) = \sum_{k=1}^{p} |x_{ik} - x_{jk}| \tag{5}$$

   Description:
   x_{ik} : data from object i on the k variable
   x_{jk} : data from object j on the k variable
   i, j : $1, 2, ..., n$; $(i \neq j_)$
   k : $1, 2, ..., p$

c. Mahalanobis Distance
   Mahalanobis distance is a distance measurement that generalizes all objects by taking into account the correlation between variables without differentiating the load of each variable

(Hair et al., 2010). According to Seber (2004), Mahalanobis distance not only can overcome the problem of differences in scale in the data, but also consider the influence of the correlation between variables. When the variables are not correlated, Mahalanobis distance can be considered the same as Euclidean distance which is standardized. Mahalanobis distance between two objects can be calculated using equation (6).

$$D_{ij}^2 = \left((\mathbf{x_i} - \mathbf{x_j})' \sum{}^{\mathbf{-1}} (\mathbf{x_i} - \mathbf{x_j}) \right) \tag{6}$$

Description:
D_{ij}^2 : square of the Mahalanobis distance of obeservation i and observation j
$\mathbf{x_i}$: coloumn vector of the observation i
$\mathbf{x_j}$: coloumn vector of the observation j
$\sum$: matrix of variances and covariances

2.1.3 Cluster Validation

Cluster validity can be used to determine the number of clusters that have been formed can explain and represent the population in general or not. Cluster validity can be used to help solve the main problem in cluster analysis, namely determining the number of clusters (group) optimum. The optimum cluster in question is the cluster that has a dense distance between objects, but has a long distance from the cluster others. The measure used is the index of Silhouettes (Hair et al., 2010). The Silhouette Index is denoted by $S(i)$ which can be calculated based on equation (7).

$$S(i) = \frac{b(i) - a(i)}{max[a(i), b(i)]} \tag{7}$$

Where:
$a(i)$: average difference of the object i with all other objects in the same group.
$b(i)$: minimum value of the average difference of object i with all objects in other groups.

The larger Silhouette index value indicates that the number of clusters optimally formed. Average $S(i)$ of all objects in a cluster shows the similarity of objects in the cluster and the accuracy of objects after grouping.

2.2 Path Analysis

Path analysis is an extension of multiple regression analysis, which can be used to determine direct or indirect relationships (Dillon & Goldstein, 1984; Abdellahi et al., 2022). The path diagram serves to describe the influence or relationship between exogenous and endogenous variables. A model usually has several related variables, and in the model, there are mediating variables (Solimun & Fernandes, 2017; **?**). Loehlin (2004) explains that a path diagram is not just a simple description of the data but also a form of a causal relationship between variables which is symbolized by the direction, as shown in the figure 1.

The path coefficient is a standardized regression coefficient, so the path coefficient can also be called the standard regression coefficient. In path analysis, transformation is carried out into standard form so that it has the same average and variance. With this standardization, the path parameters can be compared with other parameters (Hidayat & Fernandes, 2017; Sri & Solimun, 2019). According to Li (1975), data transformation is carried out so that the average becomes 0 and the variance is 1 using the following formula (8).

$$Z_{X_i} = \frac{X_i - \bar{X}}{S} \tag{8}$$

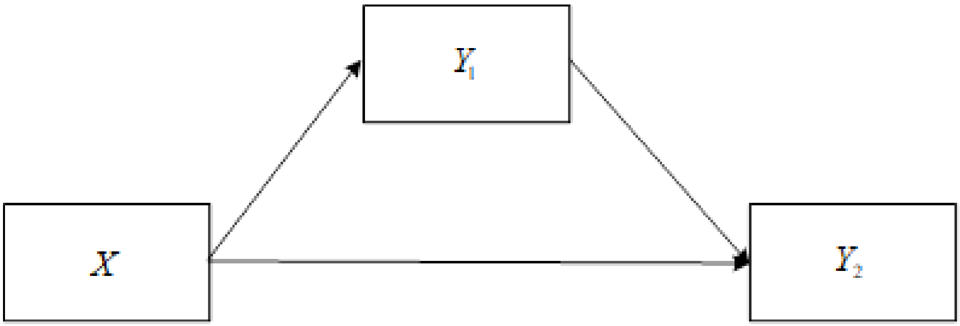

Figure 1: Simple Path Diagram (X : Exogenous Variable, Y_1: Endogenous Mediation Variable, Y_2 : Endogenous Dependent Variable)

with

$$S = \sqrt{\frac{(X_i - \bar{X})^2}{n-1}} \tag{9}$$

Where:
Z_{X_i} : exogenous variable standard score value
X_i : value of i exogenous variables
$\bar{X}$: average exogenous variables
S : standard deviation
The standard score in equation (X) is modeled by regression into the equation:

$$\begin{aligned} Z_{Y_1} &= \beta_{11} Z_{X_i} + \epsilon_{1i} \\ Z_{Y_2} &= \beta_{21} Z_{X_i} + \beta_{22} Z_{Y_{1i}} + \epsilon_{2i} \end{aligned} \tag{10}$$

The regression model in matrix form is as follows:

$$\mathbf{Z_{Y_{2n\times1}}} = \mathbf{Z_{X_{2n\times3}}} \beta_{\mathbf{3\times1}} + \epsilon_{\mathbf{2n\times1}}$$
$$\begin{bmatrix} Z_{Y11} \\ Z_{Y12} \\ Z_{Y13} \\ \vdots \\ Z_{Y1n} \\ Z_{Y21} \\ Z_{Y22} \\ Z_{Y23} \\ \vdots \\ Z_{Y2n} \end{bmatrix} = \begin{bmatrix} \mathbf{X_{Z_{XX}}} & \mathbf{0_{n\times2}} \\ \mathbf{0_{n\times1}} & \mathbf{X_{Z_{XY}}} \end{bmatrix} \begin{bmatrix} \beta_{11} \\ \beta_{21} \\ \beta_{22} \end{bmatrix} + \begin{bmatrix} \epsilon_{11} \\ \epsilon_{12} \\ \epsilon_{13} \\ \vdots \\ \epsilon_{1n} \\ \epsilon_{21} \\ \epsilon_{22} \\ \epsilon_{23} \\ \vdots \\ \epsilon_{2n} \end{bmatrix} \tag{11}$$

where:

$$\mathbf{X_{Z_{XX}}} = \begin{bmatrix} Z_{X_1} \\ Z_{X_2} \\ Z_{X_3} \\ \vdots \\ Z_{X_n} \end{bmatrix} ; \mathbf{X_{Z_{XY}}} = \begin{bmatrix} Z_{X_1} & Z_{Y_{11}} \\ Z_{X_2} & Z_{Y_{12}} \\ Z_{X_3} & Z_{Y_{13}} \\ \vdots & \vdots \\ Z_{X_n} & Z_{Y_{1n}} \end{bmatrix} \tag{12}$$

By using path analysis, causal relationships are not only analyzed directly, but also indirectly (Rutherford, 1993). The following is an explanation of each type of influence :

a. Direct Influence
   The direct effect occurs if the relationship between exogenous and endogenous variables does not require another intermediary variable. The direct effect of exogenous variables X on endogenous variables Y is β_{XY}.

b. Indirect Influence
   Indirect effect occurs when the relationship between exogenous and endogenous variables requires an intermediary of other variables. For example, the size of the exogenous variable X that influences the endogenous variable Y_2 through the endogenous variable Y_1 is $\beta_{XY_1} \times \beta_{Y_1Y_2}$.

c. Total Impact
   Total influence is the sum of direct and indirect effects.

d. Influence not analyzed
   The unanalyzed effect is the effect that arises because of the relationship between exogenous variables. For example variables X_1 and variables X_2 are correlated with each other, so the magnitude of the relationship between X_1 and Y influenced by the X_2 magnitude of the relationship $\beta_{yx_2} \times r_{x_1x_2}$, the unanalyzed effect also applies to the relationship between X_2 and Y.

e. Pseudo influence
   Pseudo-effect is an effect that arises because of a correlation between exogenous variables and more than one endogenous variable that is correlated with each other.

2.3 Dummy Regression

Drapper & Smith (1998) define regression analysis as a statistical method to explain the relationship between predictor variables and one or more response variables. In the regression model, the predictor and response variables are always quantitative. However, in its application in everyday life, there are several cases that produce regression models with qualitative predictor variables called dummy variables (Gujarati, 2003).

Regression analysis is not only used for quantitative data, but also for qualitative data obtained from the results of classifying individuals (observation units) into categories that can be ordinal or nominal. According to Gujarati (2003), dummy variable is formed as a component of the regression model in various ways :

1) Linear Regression
   Linear regression equation is given in equation (13) and Figure 2.

$$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i \tag{13}$$

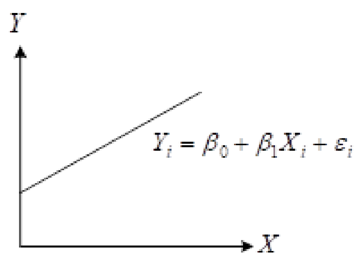

Figure 2: Linear Regression Line

2) Dummy Difference Intercept
The equation of linear regression with the intersection of dummy differences can be seen in equation (14).

$$Y_i = \beta_0 + \beta_1 X_i + \beta_1 D_i + \epsilon_i \tag{14}$$

contains the variable D which is the variable intercept, so formed two different regression equations on the components intercept:

$$Y_i = (\beta_0 + \beta_2) + \beta_1 X_i + \epsilon_i; D_i = 1$$
$$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i; D_i = 0$$

Linear regression equation with dummy difference intercepts, where $\beta_1 > 0$, and $\beta_2 > 0$ is shown in Figure 3.

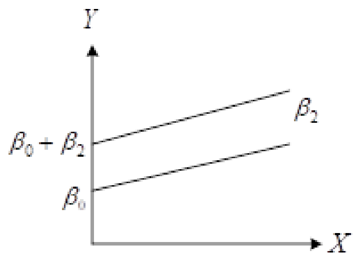

Figure 3: Linear Regression Line with Dummy Difference Intercept

3) Dummy Difference Slopes
Linear regression equation with dummy difference Slope can be seen in equation (15).

$$Y_i = \beta_{01} + \beta_{11} X_{1i} + \beta_{21} D_i X_{2i} + \epsilon_i \tag{15}$$

contains the variable D which is the slope, so that formed two different regression equations on the components slopes:

$$Y_i = \beta_0 + (\beta_1 + \beta_2) X_i + \epsilon_i; D_i = 1$$
$$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i; D_i = 0$$

Linear regression equation with dummy difference intercepts, where $\beta_1 > 0$, and $\beta_2 > 0$ is shown in Figure 4.

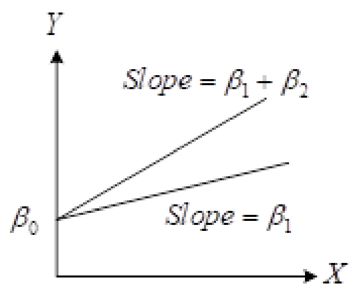

Figure 4: Linear Regression Line with Dummy Difference Slopes

4) Dummy differences Intercepts and Slopes
Linear regression equation with dummy difference intercept and slope can be seen in equation (16).

$$Y_i = \beta_0 + \beta_1 X_1 + \beta_2 D_i X_2 + \beta_3 D_i + \epsilon_i \tag{16}$$

contains the variable D which is intercept andsSlope, so that formed two different regression equations on the components intercept and slope :

$$Y_i = (\beta_0 + \beta_3) + (\beta_1 + \beta_2)X_i + \epsilon_i; D_i = 1$$
$$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i; D_i = 0$$

Linear regression equation with dummy difference intercepts and slope, where $\beta_1 > 0$, and $\beta_2 > 0$ is shown in Figure 5.

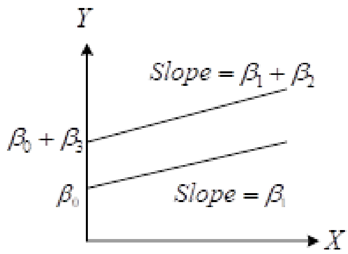

Figure 5: Linear Regression Equation with Dummy Difference Intercept And Slope

3 Materials and Methods

The research begins with the development of path analysis involving dummy variables. It is the integration of cluster analysis into the model in path analysis, which in its development is complemented by an arrangement of lemmas and proofs. Model validation is carried out using primary data from survey results.

3.1 Research Data

The data used in this study is primary data, obtained through a survey using a questionnaire with a Likert scale. A pilot test is carried out first on a questionnaire that has been formed to check the validity and reliability of the questionnaire. Then a valid and reliable questionnaire was obtained.All items used in research must be valid, because validity measures the accuracy in measuring the instruments in measuring (Solimun & Fernandes, 2017). The population in this study were customers of House Ownership Loan at Bank X. The sampling technique used was a simple random sampling of House Ownership Loan's customers at Bank X, and a sample size of $n = 200$ customers of House Ownership Loan at Bank X. The latent variable data is the indicator average score. The data obtained is used to prove and support the theoretical model within the conceptual framework used in this study (Hakim & Fernandes, 2017).

3.2 Research Model

This study uses one exogenous variable, namely attitude ($X1$) with three endogenous variables, namely behavioral intention ($Y1$), willingness to pay ($Y2$), and compliance behavior ($Y3$). Cluster analysis was conducted to group of House Ownership's customers at Bank X according to similar characteristics based on these four variables. Three distances (Euclidean, Manhattan, and Mahalanobis) and three linkages (Single, Complete, Average) are used in cluster analysis, so that cluster analysis will be carried out 9 times, which are all combinations of distances and linkages.

Theory of Planned Behavior (Ajzen, 1991) is used to design a model of the relationship between variables, which includes the variable's attitude, behavioral intention, willingness to pay, and complience behavior to pay House Ownership Behavior. The model in the form of a path diagram is presented in Figure 6. Then, cluster integration and path analysis modeling

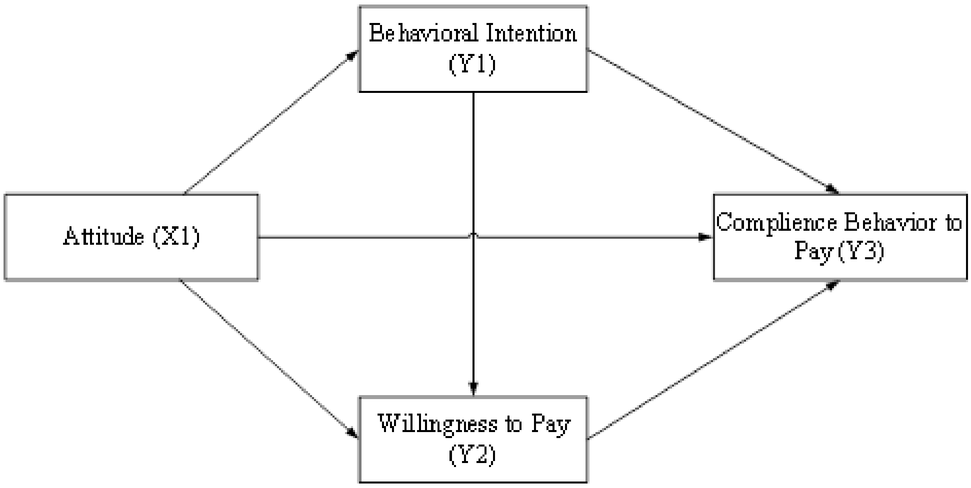

Figure 6: Research Model Diagram with Path Analysis

was formed based on dummy variables in each result of cluster analysis. Model evaluation is carried out using the adjusted total coefficient of determination (R^2 Total adjusted). From the evaluation results will obtain combination of linkage and distance has performance best for use in integrating cluster analysis and path analysis in compliance with credit payments.

The analysis process is carried out based on the results of testing assumptions both on cluster analysis assumptions also path analysis. For data that meets all assumptions, parameter estimation is performed using Ordinary Least Square (OLS). If it does not meet the normality assumption the residual is handled by resampling bootstrap. Data that does not meet the assumption of homogeneity is handled by the WLS estimation method. Then do the hypothesis testing.

4 Results

This research first presents the results of the development of a path model with a dummy variable, and secondly the results of path analysis with several distance measurement methods and grouping methods along with determining the optimum cluster. The third is the result of path analysis from the results of various combinations of methods in cluster analysis which is equipped with parameter estimation methods, adjusted total coefficient of determination, and $p-values$. The fourth is to interpret the results of the path analysis with a dummy variable on Houese Ownership Loan payment compliance behavior at Bank X.

4.1 Development of integration of Cluster Analysis and Path Modeling

Lemma 1. *Linear Path Model with Two-Category Dummy Variables*
If given paired data $(X_i, Y_{1i}, Y_{2i}, Y_{3i})$ *with* $i-1,2,3,...,n$ *following the linear path analysis model with two categories, the form of linear path analysis model with the two-category function is obtained as presented in equation* (17) *and the model in* (18).

$$\begin{aligned} Y_{1i} &= f(X_i) + \epsilon_{1i} \\ Y_{2i} &= f(X_i, Y_{1i}) + \epsilon_{2i} \\ Y_{3i} &= f(X_i, Y_{1i}, Y_{2i}) + \epsilon_{3i} \end{aligned} \tag{17}$$

Then described in the system equation

$$
\begin{aligned}
Y_{1i} &= \beta_{01} + \beta_{11}X_i + \beta_{21}D_iX_i + \epsilon_{1i} \\
Y_{2i} &= \beta_{02} + \beta_{12}X_i + \beta_{22}Y_{1i} + \beta_{32}D_iX_i + \beta_{42}D_iY_{1i} + \epsilon_{2i} \\
Y_{3i} &= \beta_{03} + \beta_{13}X_i + \beta_{23}Y_{1i} + \beta_{33}Y_{2i} + \beta_{43}D_iX_i + \beta_{53}D_iY_{1i} + \beta_{63}D_iY_{2i} + \epsilon_{3i}
\end{aligned}
\tag{18}
$$

in matrix form

$$\mathbf{Y_{3n\times 1}} = \mathbf{X_{3n\times 15}}\beta_{\mathbf{15\times 1}} + \epsilon_{\mathbf{3n\times 1}}$$

$$
\begin{bmatrix} Y_{11} \\ Y_{12} \\ \vdots \\ Y_{1n} \\ Y_{21} \\ Y_{22} \\ \vdots \\ Y_{2n} \\ Y_{31} \\ Y_{32} \\ \vdots \\ Y_{3n} \end{bmatrix}
=
\begin{bmatrix}
\mathbf{X}_{X_1DX_1} & \mathbf{0}_{n\times 5} & \mathbf{0}_{n\times 7} \\
\mathbf{0}_{n\times 3} & \mathbf{X}_{X_1Y_1DX_1DY_1} & \mathbf{0}_{n\times 7} \\
\mathbf{0}_{n\times 3} & \mathbf{0}_{n\times 5} & \mathbf{X}_{X_1Y_1Y_2DX_1DY_1DY_2}
\end{bmatrix}
\begin{bmatrix} \beta_{01} \\ \beta_{11} \\ \beta_{21} \\ \beta_{02} \\ \beta_{12} \\ \beta_{22} \\ \beta_{32} \\ \beta_{42} \\ \beta_{03} \\ \beta_{13} \\ \beta_{23} \\ \beta_{33} \\ \beta_{43} \\ \beta_{53} \\ \beta_{63} \end{bmatrix}
+
\begin{bmatrix} \epsilon_{11} \\ \epsilon_{12} \\ \vdots \\ \epsilon_{1n} \\ \epsilon_{21} \\ \epsilon_{22} \\ \vdots \\ \epsilon_{2n} \\ \epsilon_{31} \\ \epsilon_{32} \\ \vdots \\ \epsilon_{3n} \end{bmatrix}
$$

where $\mathbf{X}_{X_1DX_1}$ *is a matrix of size* $(n \times 3)$, $\mathbf{X}_{X_1Y_1DX_1DY_1}$ *is a matrix of size* $(n \times 5)$, *and* $\mathbf{X}_{X_1Y_1Y_2DX_1DY_1DY_2}$ *is a matrix of size* $(n \times 7)$ *as follows:*

$$
\mathbf{X}_{X_1DX_1} = \begin{bmatrix}
1 & X_1 & D_1X_1 \\
1 & X_2 & D_2X_2 \\
1 & X_3 & D_3X_3 \\
\vdots & \vdots & \vdots \\
1 & X_n & D_nX_{1n}
\end{bmatrix};
$$

$$
\mathbf{X}_{X_1Y_1DX_1DY_1} = \begin{bmatrix}
1 & X_1 & Y_{11} & D_1X_1 & D_1Y_{11} \\
1 & X_2 & Y_{12} & D_2X_2 & D_2Y_{12} \\
1 & X_3 & Y_{13} & D_3X_3 & D_3Y_{13} \\
\vdots & \vdots & \vdots & \vdots & \vdots \\
1 & X_n & Y_{1n} & D_nX_n & D_nY_{1n}
\end{bmatrix};
$$

$$
\mathbf{X}_{X_1Y_1Y_2DX_1DY_1DY_2} = \begin{bmatrix}
1 & X_1 & Y_{11} & Y_{21} & D_1X_1 & D_1Y_{11} & D_1Y_{21} \\
1 & X_2 & Y_{12} & Y_{22} & D_2X_2 & D_2Y_{12} & D_2Y_{22} \\
1 & X_3 & Y_{13} & Y_{23} & D_3X_3 & D_3Y_{13} & D_3Y_{23} \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
1 & X_n & Y_{1n} & Y_{2n} & D_nX_n & D_nY_{1n} & D_nY_{2n}
\end{bmatrix}
$$

where
Y_{hi} : *the* $h-th$ *endogenous variable the* $i-th$ *oveservation* $(h = 1, 2)$;$(i = 1, 2, 3, .., n)$
D_i : *the* $i-th$ *dummy variable*
X_i : *the* $i-th$ *observatioan of exogenous variable*
D_iX_i : *interaction of the* $i-th$ *dummy variable and the* $i-th$ *ecogenous variable*
n : *number of observation*
q : *number of exogenous variable*

β_{jgh}: *the* $j-th$ *coefficient of the* $g-th$ *exogenous variable and the* $h-th$ *endogenous variable* $(j=1,2,3)$;$(h=1,2)$

Proof. Before obtaining the model in simple path analysis with two categories, the model is obtained first from (a) Simple Linear Regression Analysis; (b) Simple Linear Path Analysis; and (c) Simple Linear Regression Analysis with two categories:

First part : It is known that the simple linear regression model with equation (19) and the model in (20).

$$\begin{aligned} Y_{1i} &= f(X_i) + \epsilon_{1i} \\ Y_i &= \beta_0 + \beta_1 X_i + \epsilon_i \end{aligned} \tag{19}$$

The matrix can be formed:

$$\mathbf{Y_{n\times 1}} = \mathbf{X_{n\times 2}}\beta_{\mathbf{2\times 1}} + \epsilon_{\mathbf{n\times 1}} \tag{20}$$

$$\begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} 1 & X_1 \\ 1 & X_2 \\ 1 & X_3 \\ \vdots & \vdots \\ 1 & X_n \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \\ \vdots \\ \epsilon_n \end{bmatrix}$$

where
Y_i : the $i-th$ observation of the endogenous variables $(i=1,2,3,...,n)$
X_i :tThe i-th observation exogenous variable
n : number of observations
β_j : the $j-th$ coefficient of influence of exogenous variables on endogenous $(j=0,1,2)$
ϵ_i : random error endogenous variable on the $i-th$ observation

Second part: The simple linear path analysis model is known as presented in equation (21) and the model in (22).

$$\begin{aligned} Y_{1i} &= f(X_i) + \epsilon_{1i} \\ Y_{2i} &= f(X_i, Y_{1i}) + \epsilon_{2i} \end{aligned} \tag{21}$$

Then described in the equation :

$$\begin{aligned} Y_1 i &= \beta_{01} + \beta_{11} X_i + \epsilon_{1i} \\ Y_2 i &= \beta_{02} + \beta_{12} X_i + \beta_{22} Y_{1i} + \epsilon_{2i} \end{aligned} \tag{22}$$

The matrix can be formed:

$$\mathbf{Y_{2n\times 1}} = \mathbf{X_{2n\times 5}}\beta_{\mathbf{5\times 1}} + \epsilon_{\mathbf{2n\times 1}}$$

$$\begin{bmatrix} Y_1 1 \\ Y_1 2 \\ Y_1 3 \\ \vdots \\ Y_1 n \\ Y_2 1 \\ Y_2 2 \\ Y_2 3 \\ \vdots \\ Y_2 n \end{bmatrix} = \begin{bmatrix} \mathbf{X}_X & \mathbf{0}_{n\times 3} \\ \mathbf{0}_{n\times 2} & \mathbf{X}_{XY} \end{bmatrix} \begin{bmatrix} \beta_0 1 \\ \beta_1 1 \\ \beta_0 2 \\ \beta_1 2 \\ \beta_2 2 \end{bmatrix} + \begin{bmatrix} \epsilon_1 1 \\ \epsilon_1 2 \\ \epsilon_1 3 \\ \vdots \\ \epsilon_1 n \\ \epsilon_2 1 \\ \epsilon_2 2 \\ \epsilon_2 3 \\ \vdots \\ \epsilon_2 n \end{bmatrix}$$

where

$$\mathbf{X}_X = \begin{bmatrix} 1 & X_1 \\ 1 & X_2 \\ 1 & X_3 \\ \vdots & \vdots \\ 1 & X_n \end{bmatrix};$$

$$\mathbf{X}_{XY} = \begin{bmatrix} 1 & X_1 & Y_{11} \\ 1 & X_2 & Y_{12} \\ 1 & X_3 & Y_{13} \\ \vdots & \vdots & \vdots \\ 1 & X_n & Y_{1n} \end{bmatrix}$$

with
Y_{hi} : the $h-th$ endogenous variable the $i-th$ observation $(h = 1,2)$;$i = 1,2,3,...,n$
X_i : the $i-th$ observation of exogenous variable
n : number of observations
q : number of exogenous variables
β_{jh} : the coefficient effect of the $g-th$ exogenous variable and the $j-th$ endogenous variable $(j = 0,1,2)$; $(g = 1,2)$
ϵ_{hi} : random error of the $h-th$ endogenous variable and $i-th$ observation

Third part : After knowing the equations and simple linear regression models, a simple regression model with two categories can be made as presented in equations (23) and (24).

$$Y_{1i} = f(X_i) + \epsilon_{1i} \tag{23}$$

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 D_i X_i + \epsilon_i \tag{24}$$

The matrix can be formed:

$$\mathbf{Y_{n\times 1}} = \mathbf{X_{n\times 3}}\beta_{\mathbf{3\times 1}} + \epsilon_{\mathbf{n\times 1}}$$

$$\begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} 1 & X_1 & D_1X_1 \\ 1 & X_2 & D_2X_2 \\ 1 & X_3 & D_3X_3 \\ \vdots & \vdots & \vdots \\ 1 & X_n & D_nX_n \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \\ \vdots \\ \epsilon_n \end{bmatrix}$$

where
Y_i : the $i-th$ observation of the endogenous variables $(i = 1,2,3,..,n)$
X_i : the $i-th$ observation exogenous variable to
D_i : the $i-th$ observation of the dummy variable
D_iX_i : interaction of the $i-th$ dummy variable and the $i-th$ observation exogenous variable
n : number of observations
q : number of exogenous variables
β_j : the $j-th$ coefficient of influence of exogenous variables on endogenous variable $(j = 0,1,2)$
ϵ_i : random error endogenous variable on the $i-th$ observation

From the equations in the simple linear regression analysis model, simple path analysis, and regression analysis with the two categories that have been described, we obtain a function formed as in equations (25) and (26), so that the following matrices are obtained:

$$\mathbf{Y_{3n\times 1}} = \mathbf{X_{3n\times 15}}\beta_{\mathbf{15\times 1}} + \epsilon_{\mathbf{3n\times 1}} \tag{25}$$

$$\begin{bmatrix} Y_{11} \\ Y_{12} \\ \vdots \\ Y_{1n} \\ Y_{21} \\ Y_{22} \\ \vdots \\ Y_{2n} \\ Y_{31} \\ Y_{32} \\ \vdots \\ Y_{3n} \end{bmatrix} = \begin{bmatrix} \mathbf{X}_{X_1DX_1} & \mathbf{0}_{n\times 5} & \mathbf{0}_{n\times 7} \\ \mathbf{0}_{n\times 3} & \mathbf{X}_{X_1Y_1DX_1DY_1} & \mathbf{0}_{n\times 7} \\ \mathbf{0}_{n\times 3} & \mathbf{0}_{n\times 5} & \mathbf{X}_{X_1Y_1Y_2DX_1DY_1DY_2} \end{bmatrix} \begin{bmatrix} \beta_{01} \\ \beta_{11} \\ \beta_{21} \\ \beta_{02} \\ \beta_{12} \\ \beta_{22} \\ \beta_{32} \\ \beta_{42} \\ \beta_{03} \\ \beta_{13} \\ \beta_{23} \\ \beta_{33} \\ \beta_{43} \\ \beta_{53} \\ \beta_{63} \end{bmatrix} + \begin{bmatrix} \epsilon_{11} \\ \epsilon_{12} \\ \vdots \\ \epsilon_{1n} \\ \epsilon_{21} \\ \epsilon_{22} \\ \vdots \\ \epsilon_{2n} \\ \epsilon_{31} \\ \epsilon_{32} \\ \vdots \\ \epsilon_{3n} \end{bmatrix} \tag{26}$$

where $\mathbf{X}_{X_1DX_1}$ is a matrix of size $(n \times 3)$, $\mathbf{X}_{X_1Y_1DX_1DY_1}$ is a matrix of size $(n \times 5)$, and $\mathbf{X}_{X_1Y_1Y_2DX_1DY_1DY_2}$ is a matrix of size $(n \times 7)$ as follows:

$$\mathbf{X}_{X_1DX_1} = \begin{bmatrix} 1 & X_1 & D_1X_1 \\ 1 & X_2 & D_2X_2 \\ 1 & X_3 & D_3X_3 \\ \vdots & \vdots & \vdots \\ 1 & X_{1n} & D_nX_n \end{bmatrix};$$

$$\mathbf{X}_{X_1Y_1DX_1DY_1} = \begin{bmatrix} 1 & X_1 & Y_{11} & D_1X_1 & D_1Y_{11} \\ 1 & X_2 & Y_{12} & D_2X_2 & D_2Y_{12} \\ 1 & X_3 & Y_{13} & D_3X_3 & D_3Y_{13} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & X_n & Y_{1n} & D_nX_n & D_nY_{1n} \end{bmatrix};$$

$$\mathbf{X}_{X_1Y_1Y_2DX_1DY_1DY_2} = \begin{bmatrix} 1 & X_1 & Y_{11} & Y_{21} & D_1X_1 & D_1Y_{11} & D_1Y_{21} \\ 1 & X_2 & Y_{12} & Y_{22} & D_2X_2 & D_2Y_{12} & D_2Y_{22} \\ 1 & X_3 & Y_{13} & Y_{23} & D_3X_3 & D_3Y_{13} & D_3Y_{23} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & X_n & Y_{1n} & Y_{2n} & D_nX_n & D_nY_{1n} & D_nY_{2n} \end{bmatrix}$$

□

Lemma 2. *Linear Path Model with Three-Category Dummy Variables*
If given paired data $(X_i, Y_{1i}, Y_{2i}, Y_{3i})$ *with* $i - 1, 2, 3, ..., n$ *following the linear path analysis model of categories with three categories, the form of the three-category linear path analysis function is obtained as shown in equation* (27) *and the model in* (28).

$$\begin{aligned} Y_{1i} &= f(X_i) + \epsilon_{1i} \\ Y_{2i} &= f(X_i, Y_{1i}) + \epsilon_{2i} \\ Y_{3i} &= f(X_i, Y_{1i}, Y_{2i}) + \epsilon_{3i} \end{aligned} \tag{27}$$

Then described in the system equation

$$\begin{aligned}
Y_{1i} &= \beta_{01} + \beta_{11}X_i + \beta_{21}D_{1i}X_i + \beta_{31}D_{2i}X_i + \epsilon_{1i} \\
Y_{2i} &= \beta_{02} + \beta_{12}X_i + \beta_{22}Y_{1i} + \beta_{32}D_{1i}X_i + \beta_{42}D_{1i}Y_{1i} + \beta_{52}D_{2i}X_i + \beta_{62}D_{2i}Y_{1i} + \epsilon_{2i} \\
Y_{3i} &= \beta_{03} + \beta_{13}X_i + \beta_{23}Y_{1i} + \beta_{33}Y_{2i} + \beta_{43}D_{1i}X_i + \beta_{53}D_{1i}Y_{1i} + \beta_{63}D_{1i}Y_{2i} + \beta_{73}D_{2i}X_i \\
&= +\beta_{83}D_{2i}Y_{1i} + \beta_{93}D_{2i}Y_{2i} + \epsilon_{3i}
\end{aligned} \tag{28}$$

in matrix form

$$\mathbf{Y_{3n\times 1}} = \mathbf{X_{3n\times 21}}\beta_{\mathbf{21}\times\mathbf{1}} + \epsilon_{\mathbf{3n\times 1}}$$

$$\begin{bmatrix} Y_{11} \\ Y_{12} \\ \vdots \\ Y_{1n} \\ Y_{21} \\ Y_{22} \\ \vdots \\ Y_{2n} \\ Y_{31} \\ Y_{32} \\ \vdots \\ Y_{3n} \end{bmatrix} = \begin{bmatrix} \mathbf{X}_{X_1DX_1} & \mathbf{0}_{n\times 7} & \mathbf{0}_{n\times 10} \\ \mathbf{0}_{n\times 4} & \mathbf{X}_{X_1Y_1DX_1DY_1} & \mathbf{0}_{n\times 10} \\ \mathbf{0}_{n\times 4} & \mathbf{0}_{n\times 7} & \mathbf{X}_{X_1Y_1Y_2DX_1DY_1DY_2} \end{bmatrix} \begin{bmatrix} \beta_{01} \\ \vdots \\ \beta_{21} \\ \beta_{31} \\ \beta_{02} \\ \beta_{12} \\ \vdots \\ \beta_{62} \\ \beta_{03} \\ \beta_{13} \\ \beta_{23} \\ \vdots \\ \beta_{93} \end{bmatrix} + \begin{bmatrix} \epsilon_{11} \\ \epsilon_{12} \\ \vdots \\ \epsilon_{1n} \\ \epsilon_{21} \\ \epsilon_{22} \\ \vdots \\ \epsilon_{2n} \\ \epsilon_{31} \\ \vdots \\ \epsilon_{3n} \end{bmatrix}$$

where $\mathbf{X}_{X_1DX_1}$ *is a matrix of size* $(n \times 5)$, $\mathbf{X}_{X_1Y_1DX_1DY_1}$ *is a matrix of size* $(n \times 7)$, *and* $\mathbf{X}_{X_1Y_1Y_2DX_1DY_1DY_2}$ *is a matrix of size* $(n \times 10)$ *as follows:*

$$\mathbf{X}_{X_1DX_1} = \begin{bmatrix} 1 & X_1 & D_{11}X_1 & D_{21}X_1 \\ 1 & X_2 & D_{12}X_2 & D_{22}X_2 \\ 1 & X_3 & D_{13}X_3 & D_{23}X_3 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & X_n & D_{1n}X_n & D_{2n}X_n \end{bmatrix};$$

$$\mathbf{X}_{X_1Y_1DX_1DY_1} = \begin{bmatrix} 1 & X_1 & Y_{11} & D_{11}X_1 & D_{11}Y_{11} & D_{21}X_1 & D_{21}Y_{11} \\ 1 & X_2 & Y_{12} & D_{12}X_2 & D_{12}Y_{12} & D_{21}X_2 & D_{21}Y_{12} \\ 1 & X_3 & Y_{13} & D_{13}X_3 & D_{13}Y_{13} & D_{21}X_3 & D_{21}Y_{13} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & X_n & Y_{1n} & D_{1n}X_n & D_{1n}Y_{1n} & D_{2n}X_n & D_{2n}Y_{1n} \end{bmatrix};$$

$$\mathbf{X}_{X_1Y_1Y_2DX_1DY_1DY_2} = \begin{bmatrix} 1 & X_1 & Y_{11} & Y_{21} & D_{11}X_1 & \dots & D_{31}X_1 & D_{31}Y_{11} & D_{31}Y_{21} \\ 1 & X_2 & Y_{12} & Y_{22} & D_{12}X_2 & \dots & D_{31}X_2 & D_{31}Y_{12} & D_{31}Y_{22} \\ 1 & X_3 & Y_{13} & Y_{23} & D_{13}X_3 & \dots & D_{31}X_3 & D_{31}Y_{13} & D_{31}Y_{23} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 1 & X_n & Y_{1n} & Y_{2n} & D_{1n}X_n & \dots & D_{3n}X_n & D_{3n}Y_{1n} & D_{3n}Y_{2n} \end{bmatrix}$$

where
Y_{hi} : *the* $h-th$ *endogenous variable the* $i-th$ *observation* $(h = 1, 2)$;$(i = 1, 2, 3, .., n)$
D_i : *the* $i-th$ *dummy variable*
X_i : *the* $i-th$ *observation of exogenous variable*
D_iX_i : *interaction of the* $i-th$ *dummy variable and the* $i-th$ *exogenous variable*
n : *number of observation*

q : number of exogenous variable
β_{jgh}: the $j-th$ coefficient of the $g-th$ exogenous variable and the $h-th$ endogenous variable $(j=1,2,3)$;$(h=1,2)$
$\epsilon_h i$: Random error of the h–th endogenous variable on the i–th observation

Proof. Before obtaining the model in three Category Simple Path Analysis, the model is obtained first from (a) Simple Linear Regression Analysis; (b) Simple Linear Path Analysis; and (c) Simple Linear Regression Analysis with three categories as follows:

First part : It is known that the simple linear regression model with equation (29) and the model in (30).

$$Y_{1i} = f(X_i) + \epsilon_{1i} \tag{29}$$
$$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i \tag{30}$$

The matrix can be formed:

$$\mathbf{Y_{n\times 1}} = \mathbf{X_{n\times 2}}\beta_{\mathbf{2\times 1}} + \epsilon_{\mathbf{n\times 1}}$$
$$\begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} 1 & X_1 \\ 1 & X_2 \\ 1 & X_3 \\ \vdots & \vdots \\ 1 & X_n \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \\ \vdots \\ \epsilon_n \end{bmatrix}$$

where
Y_i : the $i-th$ observation of the endogenous variables $(i = 1, 2, 3, ..., n)$
X_i : the i-th observation exogenous variable
n : number of observations
β_j : the $j-th$ coefficient of influence of exogenous variables on endogenous $(j = 0, 1, 2)$
ϵ_i : random error endogenous variable on the $i-th$ observation

Second part: The simple linear path analysis model is known as presented in equation (31) and the model in (32).

$$Y_{1i} = f(X_i) + \epsilon_{1i} \tag{31}$$
$$Y_{2i} = f(X_i, Y_{1i}) + \epsilon_{2i}$$
$$Y_1 i = \beta_{01} + \beta_{11} X_i + \epsilon_{1i} \tag{32}$$
$$Y_2 i = \beta_{02} + \beta_{12} X_i + \beta_{22} Y_{1i} + \epsilon_{2i}$$

The matrix can be formed:

$$\mathbf{Y_{2n\times 1}} = \mathbf{X_{2n\times 5}}\beta_{\mathbf{5\times 1}} + \epsilon_{\mathbf{2n\times 1}}$$
$$\begin{bmatrix} Y_1 1 \\ Y_1 2 \\ Y_1 3 \\ \vdots \\ Y_1 n \\ Y_2 1 \\ Y_2 2 \\ Y_2 3 \\ \vdots \\ Y_2 n \end{bmatrix} = \begin{bmatrix} \mathbf{X}_X & \mathbf{0}_{n\times 3} \\ \mathbf{0}_{n\times 2} & \mathbf{X}_{XY} \end{bmatrix} \begin{bmatrix} \beta_0 1 \\ \beta_1 1 \\ \beta_0 2 \\ \beta_1 2 \\ \beta_2 2 \end{bmatrix} + \begin{bmatrix} \epsilon_1 1 \\ \epsilon_1 2 \\ \epsilon_1 3 \\ \vdots \\ \epsilon_1 n \\ \epsilon_2 1 \\ \epsilon_2 2 \\ \epsilon_2 3 \\ \vdots \\ \epsilon_2 n \end{bmatrix}$$

where

$$\mathbf{X}_X = \begin{bmatrix} 1 & X_1 \\ 1 & X_2 \\ 1 & X_3 \\ \vdots & \vdots \\ 1 & X_n \end{bmatrix};$$

$$\mathbf{X}_{XY} = \begin{bmatrix} 1 & X_1 & Y_{11} \\ 1 & X_2 & Y_{12} \\ 1 & X_3 & Y_{13} \\ \vdots & \vdots & \vdots \\ 1 & X_n & Y_{1n} \end{bmatrix}$$

with
Y_{hi} : the $h-th$ endogenous variable the $i-th$ observation $(h=1,2)$;$i=1,2,3,...,n$
X_i : the $i-th$ bservation of exogenous variable
n : number of observations
q : number of exogenous variables
β_{jh} : the coefficient effect of the $g-th$ exogenous variable and the j-th endogenous variable $(j=0,1,2)$; $(g=1,2)$
ϵ_{hi} : random error of the $h-th$ endogenous variable and $i-th$ observation

Third part : After knowing the equations and simple linear regression models, a simple regression model with two categories can be made as presented in equations (33) and (34).

$$Y_{1i} = f(X_i) + \epsilon_{1i} \tag{33}$$

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 D_{1i} X_i + \beta_3 D_{2i} X_i + \epsilon_i \tag{34}$$

The matrix can be formed:

$$\mathbf{Y_{n\times 1}} = \mathbf{X_{n\times 4}} \beta_{\mathbf{4\times 1}} + \epsilon_{\mathbf{n\times 1}}$$

$$\begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} 1 & X_1 & D_{11}X_1 & D_{21}X_1 \\ 1 & X_2 & D_{12}X_2 & D_{22}X_2 \\ 1 & X_3 & D_{13}X_3 & D_{23}X_3 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & X_n & D_{1n}X_n & D_{2n}X_n \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_2 3 \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \\ \vdots \\ \epsilon_n \end{bmatrix}$$

where
Y_i : the $i-th$ observation of the endogenous variables $(i=1,2,3,..,n)$
X_i : the $i-th$ observation exogenous variable to
D_i : the $i-th$ observation of the dummy variable
$D_i X_i$: interaction of the $i-th$ dummy variable and the $i-th$ observation exogenous variable
n : number of observations
q : number of exogenous variables
β_j : the $j-th$ coefficient of influence of exogenous variables on endogenous variable $(j=0,1,2)$
ϵ_i : random error endogenous variable on the $i-th$ observation

From the equations in the simple linear regression analysis model, simple path analysis, and regression analysis with the three categories. As explained above, we can obtain the function formed as in equations (35) and (36), so that the following matrices are obtained:

$$\mathbf{Y_{3n\times 1}} = \mathbf{X_{3n\times 21}}\beta_{\mathbf{21\times 1}} + \epsilon_{\mathbf{3n\times 1}} \tag{35}$$

$$\begin{bmatrix} Y_{11} \\ Y_{12} \\ \vdots \\ Y_{1n} \\ Y_{21} \\ Y_{22} \\ \vdots \\ Y_{2n} \\ Y_{31} \\ Y_{32} \\ \vdots \\ Y_{3n} \end{bmatrix} = \begin{bmatrix} \mathbf{X}_{X_1DX_1} & \mathbf{0}_{n\times 7} & \mathbf{0}_{n\times 10} \\ \mathbf{0}_{n\times 4} & \mathbf{X}_{X_1Y_1DX_1DY_1} & \mathbf{0}_{n\times 10} \\ \mathbf{0}_{n\times 4} & \mathbf{0}_{n\times 7} & \mathbf{X}_{X_1Y_1Y_2DX_1DY_1DY_2} \end{bmatrix} \begin{bmatrix} \beta_{01} \\ \vdots \\ \beta_{21} \\ \beta_{31} \\ \beta_{02} \\ \beta_{12} \\ \vdots \\ \beta_{62} \\ \beta_{03} \\ \beta_{13} \\ \beta_{23} \\ \vdots \\ \beta_{93} \end{bmatrix} + \begin{bmatrix} \epsilon_{11} \\ \epsilon_{12} \\ \vdots \\ \epsilon_{1n} \\ \epsilon_{21} \\ \epsilon_{22} \\ \vdots \\ \epsilon_{2n} \\ \epsilon_{31} \\ \vdots \\ \epsilon_{3n} \end{bmatrix} \tag{36}$$

where $\mathbf{X}_{X_1DX_1}$ is a matrix of size $(n \times 5)$, $\mathbf{X}_{X_1Y_1DX_1DY_1}$ is a matrix of size $(n \times 7)$, and $\mathbf{X}_{X_1Y_1Y_2DX_1DY_1DY_2}$ is a matrix of size $(n \times 10)$ as follows:

$$\mathbf{X}_{X_1DX_1} = \begin{bmatrix} 1 & X_1 & D_{11}X_1 & D_{21}X_1 \\ 1 & X_2 & D_{12}X_2 & D_{22}X_2 \\ 1 & X_3 & D_{13}X_3 & D_{23}X_3 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & X_n & D_{1n}X_n & D_{2n}X_n \end{bmatrix};$$

$$\mathbf{X}_{X_1Y_1DX_1DY_1} = \begin{bmatrix} 1 & X_1 & Y_{11} & D_{11}X_1 & D_{11}Y_{11} & D_{21}X_1 & D_{21}Y_{11} \\ 1 & X_2 & Y_{12} & D_{12}X_2 & D_{12}Y_{12} & D_{21}X_2 & D_{21}Y_{12} \\ 1 & X_3 & Y_{13} & D_{13}X_3 & D_{13}Y_{13} & D_{21}X_3 & D_{21}Y_{13} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & X_n & Y_{1n} & D_{1n}X_n & D_{1n}Y_{1n} & D_{2n}X_n & D_{2n}Y_{1n} \end{bmatrix};$$

$$\mathbf{X}_{X_1Y_1Y_2DX_1DY_1DY_2} = \begin{bmatrix} 1 & X_1 & Y_{11} & Y_{21} & D_{11}X_1 & \dots & D_{31}X_1 & D_{31}Y_{11} & D_{31}Y_{21} \\ 1 & X_2 & Y_{12} & Y_{22} & D_{12}X_2 & \dots & D_{31}X_2 & D_{31}Y_{12} & D_{31}Y_{22} \\ 1 & X_3 & Y_{13} & Y_{23} & D_{13}X_3 & \dots & D_{31}X_3 & D_{31}Y_{13} & D_{31}Y_{23} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 1 & X_n & Y_{1n} & Y_{2n} & D_{1n}X_n & \dots & D_{3n}X_n & D_{3n}Y_{1n} & D_{3n}Y_{2n} \end{bmatrix}$$

□

Integration of cluster analysis and path analysis that produce a path model involving the dummy variable, which is then implemented to analyze the complience behavior to pay credit to bank X.

4.2 Selection of the best Cluster Analysis Results from a Combination of Distance and Linkage

This study uses three types of distance, namely Euclidean, Manhattan, and Mahalanobis, combined with three types of linkage, namely Single Linkage, Complete Linkage, and Average Linkage. The data in this study fulfill the assumption of non-multicollinearity so that it can proceed to cluster analysis with various distances. The results of cluster analysis using a combination of three distances and three linkages will be evaluated using the Silhouette index. The summary of the results of calculating the Silhouette index of all combinations is presented in Table 1.

Table 1: The Goodness of The Results of Cluster Analysis on Each Cluster Method Combination based on the Silhouette Index

Distance	Linkages	Number of Clusters	Shilhoutte
Euclidean	Single	2	0.144
	Complete	2	0.095
	Average	3	0.089
Manhattan	Single	2	0.228
	Complete	2	0.312
	Average	3	0.143
Mahalanobis	Single	2	0.298
	Complete	**2**	**0.365**
	Average	2	0.295

By using the Silhouette index, the optimal number of clusters is selected based on the largest value index. Based on Table 1, the optimal number of clusters selected is 2 clusters. It can be seen that the most optimal combination of methods is the combination of Mahalanobis distance and Complete Linkage. The analysis using a combination of the cluster method shows better results at grouping House Ownership Loan's customers at Bank X to other combinations of cluster methods. Thus, modeling the integration of cluster analysis with path analysis with dummy variables is formed by involving two categories.

4.3 Results of Cluster Integration Analysis with Path Analysis Using Dummy Variables

The results of cluster analysis are used as an input in the path analysis, as a dummy variable. modeling Cluster integration and path analysis were carried out using dummy variables to accommodate group status creditors. Based on the optimal cluster results obtained two clusters, then used one dummy variable ($D = 0$ for cluster 1 and $D = 1$ for cluster 2).

Cluster integration with path analysis is carried out by handling assumption violations according to the assumption test results obtained. All path models through a combination of methods in path analysis have met the assumption of linearity. Because the data meets the assumption of linearity, it can be used modeling with parametric methods (Fernandes et al, 2019). For data that meets all assumptions, parameter estimation is performed using Ordinary Least Square (OLS). Data that does not meet the residual normality assumption is handled by bootstrap resampling. If the data does not meet the residual homogeneity assumption it is handled by using WLS as a parameter estimation method.

The following presents the results of evaluating the effect of each variable on modeling with one of the combinations (Mahalanobis Distance and Complete Linkage) in Table 2.

Based on Table 2 it can be seen that in modeling using the Mahalanobis-Complete combination, several path coefficients are not significant. As many as 75% of coefficients have significance in the models. Then, in the same way, it is calculated percentage of significant coefficients, then ranked and presented in Table 4.

Evaluation of the model is seen using the $R^2$2 Total adjusted. The R^2 Total adjusted describes the total variance of data that can be represented by the model. The R^2 Total adjusted own range mark between 0% to 100%. The results of the evaluation of the cluster integration model and path analysis with dummy variables through the combination of cluster analysis are presented in Table 3. The selection of the best model is seen from the R^2 Total adjusted, based on Table 3 it can be seen that the cluster analysis integration model with path analysis with a combination of Mahalanobis distance and Complete Linkage. This means that the variables of attitude, behavior intention, and willingness to pay are able to explain the variation in the variable of complience behavior to pay by 86.63%, while the other 13.37% is explained by variables outside the model. Cluster integration using Mahalanobis distance and

Table 2: Path Coefficient Hypothesis Testing on Integration of Cluster Analysis on Mahalanobis Distance and Complete Linkage with Path Analysis

Relationship Between Variables	**Path Coefficient**	**p – value**	**Result**
$Z_{X1} \to Z_{Y1}$	0.376	0.002	significant
$Z_{X1} \to Z_{Y2}$	0.189	0.034	significant
$Z_{Y1} \to Z_{Y2}$	0.414	0.001	significant
$Z_{X1} \to Z_{Y3}$	0.375	0.024	significant
$Z_{Y1} \to Z_{Y3}$	0.187	0.023	significant
$Z_{Y2} \to Z_{Y3}$	0.443	0.001	significant
$Z_{X1}D_1 \to Z_{Y1}$	0.010	0.461	not significant
$Z_{X1}D_1 \to Z_{Y2}$	0.127	0.041	significant
$Z_{Y1}D_1 \to Z_{Y2}$	0.068	0.423	not significant
$Z_{X1}D_1 \to Z_{Y3}$	-0.178	0.094	not significant
$Z_{Y1}D_1 \to Z_{Y3}$	0.245	0.021	significant
$Z_{Y2}D_1 \to Z_{Y3}$	-0.165	0.011	significant
Percentage Significant Coefficient			75%

Table 3: Adjusted Total Coefficient of Determination on Integration of Cluster and Path analysis

Distance	**Linkages**	**R^2 Total Adjusted**
Euclidean	Single	0.7162
	Complete	0.6781
	Average	0.6623
Manhattan	Single	0.7210
	Complete	0.8124
	Average	0.7097
Mahalanobis	Single	8082
	Complete	**0.8663**
	Average	0.8052

complete linkage with path analysis is the best method used in modeling the compliance behavior to pay House Ownership Loan at Bank X. Evaluation of the integration results of cluster analysis and path analysis for each combination is summarized in Table 4.

Table 4: The Ranking of the Goodness of the Integration Model is based on the Goodness of Cluster Analysis and Path Modeling

Cluster Analysis			**Path Analysis**		
Distance	**Linkages**	**Goodness Cluster Silhoutte**	**Parameter Estimation Method**	**R^2 Total Adjusted**	**Percentage Significant Relationship**
Euclidean	Single	0.144 (6)	WLS-Resampling	0.7162 (6)	75.00% (1)
	Complete	0.095 (8)	WLS	0.6781 (8)	66.67% (2)
	Average	0.089 (9)	WLS	0.6623 (9)	66.67% (2)
Manhattan	Single	0.228 (5)	WLS-Resampling	0.7210 (5)	66.67% (2)
	Complete	0.312 (2)	OLS-Resampling	0.8124 (2)	58.33% (3)
	Average	0.143 (7)	OLS-Resampling	0.7097 (7)	75.00% (1)
Mahalanobis	Single	0.298 (3)	OLS	0.8082 (3)	75.00% (1)
	Complete	**0.365 (1)**	**OLS**	**0.8663 (1)**	**75.00% (1)**
	Average	0.295 (4)	OLS	0.8052 (4)	66.67% (2)

Note: numbers in parentheses are ratings (lower values indicate better ratings)

From Table 4 it can be seen that based on the two criteria of model goodness (cluster and path analysis), the best combination of distance and linkage is found in the Mahalanobis-Complete linkage. This shows that the criteria for the goodness of the cluster-path model combination are seen based on the Silhouette index, R^2 Total adjusted and hypothesis testing tends to be consistent.

4.4 Results of Integration of Cluster Analysis and Path Analysis about Compliance Behavior to Pay in Bank X

Complete linkage method with Mahalanobis distance as a method produces 2 clusters that separate each data set optimally. Then a path model is formed from the Integration of Cluster Analysis and Path Analysis on the payment compliance behavior of House Ownership Loan at Bank X presented in the following system of equations.

$$
\begin{aligned}
Z_{Y_1} &= 0.376Z_{X_{1i}} + 0.010D_{1i}Z_{X_{1i}}\\
Z_{Y_2} &= 0.189Z_{X_{1i}} + 0.414Z_{Y_{1i}} + 0.127D_{1i}Z_{X_{1i}} + 0.068D_{1i}Z_{Y_{1i}}\\
Z_{Y_3} &= 0.274Z_{X_{1i}} + 0.187Z_{Y_{1i}} + 0.443Z_{Y_{2i}} - 0.178D_{1i}Z_{X_{1i}} + 0.245D_{1i}Z_{Y_{1i}} + 0.165D_{1i}Z_{Y_{2i}}
\end{aligned}
$$

Clusters ' Low ' can be seen in the following ($D = 0$).

$$
\begin{aligned}
Z_{Y_1} &= 0.376Z_{X_{1i}}\\
Z_{Y_2} &= 0.189Z_{X_{1i}} + 0.414Z_{Y_{1i}}\\
Z_{Y_3} &= 0.274Z_{X_{1i}} + 0.187Z_{Y_{1i}} + 0.443Z_{Y_{2i}}
\end{aligned}
$$

Clusters ' High ' can be seen in the following equation ($D = 1$).

$$
\begin{aligned}
Z_{Y_1} &= 0.386Z_{X_{1i}}\\
Z_{Y_2} &= 0.306Z_{X_{1i}} + 0.482Z_{Y_{1i}}\\
Z_{Y_3} &= 0.197Z_{X_{1i}} + 0.432Z_{Y_{1i}} + 0.278Z_{Y_{2i}}
\end{aligned}
$$

The behavior of the relationship between variables in each cluster can be seen from the magnitude of the path coefficient in the system equation for high clusters ($D = 1$) and low clusters ($D = 0$). From the system of equations it can be seen that in the low cluster ($D = 0$), attitude (X) has a stronger effect on willingness to pay ($Y2$), as well as the influence of willingness to pay ($Y2$) on compliance behavior to pay House Ownership Loan ($Y3$). Whereas in the high cluster ($D = 1$), Attitude (X) has a stronger influence on Behavioral Intention ($Y1$) and compliance behavior to pay ($Y3$). In addition, behavioral intention ($Y1$) has a stronger effect on willingness to pay ($Y2$) and compliance behavior to pay House Ownership Loan ($Y3$).

No specific relationship patterns were obtained from the equations in the two clusters, so that the integrated cluster and path analysis was packaged in a path equation model with dummy variables, resulting in different path coefficients in the different clusters. If this difference is tested, it can be seen as a development from the analysis of multigroup moderating variables on path analysis. This is in accordance with several opinions (Heredia-Rojas et al, 2022; Pico-Saltos et al., 2023). Multi-group moderation analysis was carried out on moderating variables with categorical data (low, high) and the regression coefficients or paths were tested for differences in the two groups (categories).

5 Conclusion

The conclusions obtained are based on the results of the analysis.

1) In data with a relatively large sample size, there are groups of behavioral relationships between variables. So that the application of path analysis needs to be integrated with cluster analysis, through the estimation and testing of path models involving dummy variables. This study can be seen as the development of a multigroup moderating variable in path analysis.

2) Integration of cluster analysis with path modeling, there was no consistency of the Silhouette Index, adjusted total coefficient of determination and parameter hypothesis testing from various combinations of cluster methods studied. However, there is the best integration of cluster analysis and path modeling, which has the consistency of the goodness of the model in terms of the Silhouette index, adjusted total coefficient of determination and parameter hypothesis testing. The integration of cluster analysis with the best path modeling is the Mahalanobis-Complete Linkage combination, and a path model with an adjusted total coefficient of determination of 0.8663 is obtained, so that the model is considered good in explaining the behavior of House Ownership Loan payment compliance at the Bank X.

3) The best integration of cluster analysis and path modeling results that in low clusters ($D = 0$), attitude has a stronger effect on willingness to pay, as well as the effect of willingness to pay on compliance paying behavior. Whereas in the high cluster ($D = 1$), attitude has a stronger effect on behavioral intention and paying compliance behavior. In addition, behavioral intentions have a stronger effect on willingness to pay and compliance behavior in paying House Ownership Loan at Bank X.

6 Acknowledgement

This research received special funding from the Government of Indonesia, especially the Directorate General of Higher Education, Research and Technology and the Ministry of Education, Culture, Research and Technology. Thank you to all parties for their support and input in the preparations of this research. All the support and inputs given are very useful for the perfection of this research.

7 Conflict of Interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

Abdellahi, S.A.E., M'barek, I., & El Hadri, Z. (2022). Generalizing the properties of Finite Iterative Method for the Computation of the Covariance Matrix Implied by a Recursive Path Model. *Statistics, Optimization & Information Computing*, *10*(4), 1222-1234.

Afifi, A.A., Clark, V. (1990). *Computer-Aided Multivariate Analysis. 2nd Ed.*. New York: Nostrand Reinhold Co.

Agama, F.T., Repalle, V.S. (2022). A Study on an Extended Total Fuzzy Graph. *Appl. Math.*, *16*(4).

Ajzen. I. (1991). Theory of Planned Behavior. *Organizational Behavior and Human Decision Processes*, *50*(2).

Alharbi, M.G., Khalifa, H.A. (2021). On solutions of fully fuzzy linear fractional programming problems using close interval approximation for normalized heptagonal fuzzy numbers. *Appl. Math. Inf. Sci.*, *15*(4).

Dicky, T. (2016). *Classification of Chemical Items with Cluster Analysis at Pt Pupuk Kalimantan Timur.* (Dictoral dissertation, UAJY).

Dillon, W.R., Goldsetein, M. (1984). *Multivariate Analysis: Method and Application.* Canada: John Wiley & Sons, Inc.

Drapper, N.R., Smith, H.(1998). *Multivariate Analysis: method and Application.* Canada: John Wiley & Sons, Inc.

El-Sanowsy, E., Atef, A. (2022). (R,S)-Fuzzy G*P-Closed Sets and its Applications. *Appl. Math. Inf. Sci.*, *16*(1).

Ersoy,B., Onar, S., Hila, K., & Naka, K. (2021). (T, S)-Intuitionistic Fuzzy Hyperideals of Tau-Hyperrings, *Applied Mathematics and Information Sciences*, *15*(3).

Fernandes, A.A.R., Taba, I.M. (2019). Welding technology as the moderation variable in the relationships between government policy and quality of human resources and workforce competitiveness. *Journal of Science and Technology Policy Management*, *10*(1), 58-72.

Fernandes, A.A.R., Hutahayan, B., Arisoesilaningsih, E., Yanti, I., Astuti, A.B., & Amaliana, L.(2019). Comparison of Curve Estimation of the Smoothing Spline Nonparametric Function Path based on PLS and PWLS in Various Levels of Heteroscedasticity. *In IOP Conference Series: Materials Science and Engineering,546*(5).

Fernandes, A.A.R., Solimun, N., & Hutahayan, B. (2020). Comparison of use of linkage in integrated cluster with discriminal analysis approach. *International Journal of Advanced Science and Technology*, *29*(3), 5654-5668.

Gujarati, D.N. (2003). *Basic Econometrics.* New York: Mc.Graw-Hill.

Hair, J.F., Black, W.C., Babin, B.J., & Anderson, R.E. (2010). *Multivariate Data Analysis. 7th Ed.* New Jersey: Pearson Prentice Hall.

Hakim, W., Fernandes, A.A.R. (2017). Moderation effect of Organizational Citizenship Behavior on The Performance of lecturers. *journal of Organizational Change Management*, *30*(7).

Heredia-Rojas, B., Li Liu., Irez-Correa, P., & Mariano-melo, A. (2022). Moderating Effects of Requirements Uncertainty and Project Complexity on Value Creation Processes and Project Value: A Multigroup Analysis via Partial Least Square Structural Equation Modeling. Hindawi *Mathematical Problem in Engineering.*

Hidayat, M.F., Fernandes, A.A.R. (2019). Estimation of Truncated Spline Function in Nonparametric Path Analysis Based on Weighted Least Square (WLS). *In IOP Conference Series: Materials Science and Engineering*, *546*(5).

Jain, A.K., Dubes, R.C. (1988). *Algorithms for clustering data.* New Jersey: Pearson Prentice Hall.

Johnson, R.A. & Wichern, D.W. (1992). *Applied Multivariate Statistical Analysis. 3rd Ed..* New Jersey: Pearson Prentice Hall.

Johnson, R.A., Wichern, D.W. (2002). *Applied Multivariate Statistical Analysis. 5td Ed..* New Jersey: Pearson Prentice Hall.

Khalifa, H.A., Kumar, P., & Hassan, B. (2021). An inexact rough interval of normalized heptagonal fuzzy numbers for solving vendor selection problem. *Applied Mathematics and Information Sciences*, *15*(3), 317-324.

Kirono, I., Armanu, A., Hadiwidjojo, D., & Solimun. (2019). Logistics performance collaboration strategy and information sharing with logistics capability as mediator variable (study in Gafeksi East Java Indonesia). *International Journal of Quality & Reliability Management*, *36*(8), 1301-1317.

Kozae, A.M., Shokry, M., & Omran, M. (2021). Comparison Between Fuzzy Soft Expert System and Intuitionistic Fuzzy Set in Prediction of Luge Cancer. *Information Sciences Letters*, *10*(2).

Lashin, M., Malibari, A. (2022). Using fuzzy logic control system as an artificial intelligence tool to design soap bubbles robot as a type of interactive games. *Information Sciences Letters*, *11*(1), 15-19.

Li, C.C. (2004). *Path Analysis-a primer*. California: The Boxwood Press.

Loehlin, J.C. (2004). *Latent Variable Models: An Introducton to Factor, Path, and Structural Analysis. 4th Ed.*. New Jersey: Lawrence Erlbaum Associates, Inc.

Lone, M.A., Mir, S.A., Al-Bayatti, H. M., Özer, O., Khan, O.F.,& Mushtaq, T. (2021). Optimal allocation in agriculture using intuitionistic fuzzy assignment problem. *Information Sciences Letters*.

Marsili, F., Bödefeld, J. (2021). Integrating Cluster Analysis into Multi-Criteria Decision Making for Maintenance Management of Aging Culverts. *Mathematics*, *9*(20), 2549.

Pico-Saltos, R., Sabando-Vera, D., Yonfa-Medranda, M., & Garzas, J. (2023). Hierarchical Component Model (HCM) of Career Success and the Moderating Effect of Gender. from the Perspective of University Alumni: Multigroup Analysis and Empirical Evidence from Quevedo, Ecuador. *Sustainability 2023*, *15*(30).

Rumahorbo, A.S., Suhada, Solimun, & Hidayat, K. (2022). The improvement of firm performance and sustainability: militancy, customer focus, and bank agent experience as a moderating agent. *International Journal of Business Innovation and Research*, *28*(4). 542-560.

Rutherford, R. (1993). *Statistical Model For Causal Analysis*. New York: John Wiley & Sons, Inc.

Seber, G.A.F. (2004). *Multivariate Observations*. New York: John Wiley & Sons, Inc.

Sharma, S. (1996). *Applied Multivariate Techniques*. New York: John Wiley & Sons, Inc.

Siregar, M.F.A.N., Hakim, A., Mardiyono, & Solimun. (2022). Utility on Etle and Trust on Etle as A Booster of Social Order and Habit Through Traffic Compliance and Awareness Based on Information Technology. *Journal of Theoretical and Applied Information Technology*, *100*(4), 1027-1037

Solimun, Fernandes, A.A.R. (2017). Investigation the Mediating Variable: What is necessary?(Case study in management research). *International Journal of Law and Management*, *59*(6), 1059-1067.

Solimun, Fernandes, A.A.R. (2017). Investigation of instrument validity: Investigate the consistency between criterion and unidimensional in instrument validity (Case study in management research). *International Journal of Law and Management*, *59*(6), 1203-1210.

Sri, M., Solimun (2019). The relationship between audit quality and risk taking toward value creation in Indonesia. *Journal of Accounting in Emerging Economies*, *9*(2), 1-267.

Sumardi, S., Fernandes, A.A.R. (2020). The influence of quality management on organization performance: service quality and product characteristics as a medium. *Property Management*, *38*(3), 383-403

Wierzchoń, S.T. Ktopetek M.A. (2018). *Modern Algorithms of Cluster Analysis*. Switzerland: Springer International Publishing AG.

Wijayanto, S.A., Suhadak, S., Mangesti, S., Hidayat, K., & Solimun. (2021). Transglobal Leadership Toward Sustainability: A Case Study in Indonesia. *The Journal of Asian Finance, Economics and Business*, *8*(2). 1181-1188.