第 48 卷 第 4 期
2021 年 4 月

湖南大学学报（自然科学版）
Journal of Hunan University（Natural Sciences）

Vol. 48. No. 4.
Apr. 2021

Open Access Article

Comparison of Cluster and Linkage Validity Indices in Integrated Cluster Analysis with *Structural Equation Modeling* War-PLS Approach

Adji Achmad Rinaldo Fernandes

Faculty of Math and Science, Brawijaya University, Malang, Indonesia

Abstract. This study wants to compare the Integrated Cluster and SEM models of the Warp-PLS approach with various cluster and linkage validity indices on Service Quality, Environment, Fashions, Willingness to Pay, and Compliant Behavior of Bank X Customer Paying Behavior. The data used in this study are primary. The variables used in this study are service quality, environment, fashion, willingness to pay, and compliance with paying behavior at Bank X. The data were obtained through a questionnaire with a Likert scale — the measurement of variables in primary data using the average score of each item. The sampling technique used was purposive sampling. The object of observation is the customer as many as 100 respondents. Data analysis was carried out quantitatively; the descriptive study was carried out first, then Integrated Cluster analysis and SEM analysis of the Warp-PLS approach were carried out with euclidean distances on 4 cluster validity indices, including Index Sillhouette, Krzanowski-Lai, Dunn, Davies-Bouldin, as well as on various linkages (Ward, Average, and Complete Linkage). This research uses R software. The results show that the Silhouette, Krzanowski-Lai, Dunn, and Davies-Bouldin indexes, the complete linkage method is better than the ward and average linkage methods. The novelty in this research is Integrated Cluster Analysis and SEM of the Warp-PLS approach to compare 4 cluster validity indices and three linkages.

Keywords: Cluster Analysis; SEM; Warp-PLS; Integration Model; Dummy Variable; Cluster Validity Index; Linkage

使用結構方程模型進行集成聚類分析中的聚類和鏈接有效性指標的比較 War-PLS 方法

摘要：本研究旨在將 Warp-PLS 方法的集成聚類和 SEM 模型與各種聚類和鏈接有效性指標（服務質量，環境，時尚，支付意願以及 X 銀行客戶支付行為的合規行為）進行比較。這項研究是主要數據。本研究中使用的變量是服務質量，環境，方式，支付意願以及在 X 銀行的支付行為。數據通過利克特量表通過問卷調查獲得。使用每個項目的平均得分來測量主要數據中的變量。使用的採樣技術是有目的的採樣。觀察對像是多達 100 位受訪者的客戶。對數據進行定量分析，首先進行描述性分析，然後使用歐氏距離對 4 個聚類有效性指標進行 Warp-PLS 方法的綜合聚類分析和 SEM 分析，包括：Sillhouette 指數，Krzanowski-Lai，Dunn 指數，戴維斯・布爾丁，以及各種鏈接（病房，平均和完全鏈接）。本研究使用 R 軟件，結果表明 Silhouette，Krzanowski-Lai，Dunn 和 Davies-Bouldin 索引的完全鏈接方法優於病房和平均鏈接方法。 Warp-PLS 方法的分析和 SEM 以比較 4 個聚類有效性指標和 3 個關聯。

關鍵字：聚類分析; 掃描電鏡 經紗-PLS; 整合模型；虛擬變量; 聚類有效性指數；連鎖

1. Introduction

The existence of a bank is essential for society, one of which is shown by the increasing number of customers. There are several service products in the banking business, and one of them is a Purchasing Home Loan (KPR). KPR is one of the credit services provided by banks to customers who apply for particular loans to meet the need for a residence. Before a bank gives credit to a debtor, it is necessary to assess the bank to measure whether the debtor can pay his obligations in honor or not. One of the problems in KPR is debtors who are not smooth in paying their obligations. It causes losses to the bank. Therefore, it is necessary to have supervision in customer credit services. Namely, the classification is used to determine the characteristics of debtors who can meet credit obligations or not. Statistical methods that can be

Received: 16 February 2021 / Revised: 13 March 2021 / Accepted: 17 March 2021 / Published: 30 April 2021
About the author: Adji Achmad Rinaldo Fernandes, Faculty of Math and Science, Brawijaya University, Malang, Indonesia

used to deal with these problems are using cluster integration and path analysis.

Cluster analysis is one of the multiple variable (multivariate) analyses included in the interdependency method; namely, the independent or explanatory variables are not differentiated from the dependent variable or response. Cluster analysis aims to classify objects into several clusters, where clusters have different properties. According to Solimun [1], Cluster analysis is used to group homogeneous objects into one and inter-group characteristics, namely heterogeneous. In general, there are two methods in cluster analysis, namely the hierarchical method and the non-hierarchical method. The hierarchical approach consists of several methods: the Single Linkage method, the Average Linkage method, the Complete Linkage method, the Centroid Linkage method, and the Ward method (Ward's Method). The method that is included in the non-hierarchical approach is the K-Means method. Using cluster analysis will obtain the best number of clusters from several linkages and the cluster validity index. Then, one of the best clusters was selected for further research. In cluster analysis, several links can be used to form clusters. According to Supranto [2], the linkage method consists of Ward linkage, Complete linkage, and Average linkage. In cluster analysis, several links can be used to form clusters.

Apart from using cluster analysis, this study also uses the SEM approach of the Warp-PLS approach. SEM is a development of Path Analysis which Wright developed in 1934. Path analysis is also known as a causal model because path analysis allows theoretical testing proportions regarding certain variables' cause and effect relationship. Path analysis ([3], [4]) has weaknesses. Namely, the relationship between variables cannot be reciprocal (back and forth arrows). The model is not flexible, considering only the form of linear relationships can be accommodated in the model and is not rigid with the assumption of normality and homoscedasticity.

Research related to the integration of Clusters with SEM with the Warp-PLS approach is what Liu et al. [5] entitled "An improved path-based Clustering algorithm" shows that path analysis with Cluster integration provides results that outperform other Cluster algorithms. This study examines the integrated Cluster in SEM with the Warp-PLS approach with three different linkage methods: Ward linkage, Complete linkage, and Average linkage and four different cluster validity indices Sillhouette Index, Krzanowski-Lai, Dunn, Davies-Bouldin. Measurement of the distance of each link in this study using the Euclidean distance. The result of determining many clusters with different linkage will undoubtedly give different results. In this research, we want to study the effect of using the link on an integrated cluster with SEM with the Warp-PLS approach. This research intends to get the best linkage and cluster validity index used in an integrated cluster with SEM with the Warp-PLS approach to classifying the compliant behavior of paying KPR type loans.

In this study, the researcher will compare the integrated cluster analysis model and the SEM approach of the Warp-PLS approach using the Euclidean distance measure and different cluster and linkage validity indices. Therefore, the cluster distance size will be compared with four cluster validity indices: Silhouette, Krzanowski-Lai, Dunn, Davies-Bouldin, and three linkages, namely Ward, Average, and Complete Linkage. The cluster validity test is used to evaluate the results of the quantitative cluster analysis so that the optimum group is produced. An optimum group is a group that has a dense distance between individuals in the group and is well isolated from other groups [6]. The originality of this research is the application of Integrated Cluster Analysis and SEM of the Warp-PLS approach to compare 4 cluster validity indices and three linkages where no study discusses the comparison of 4 cluster validity indices and three links.

2. Literature Review

2.1 Analysis Cluster

Cluster analysis is a method in multiple variable analysis that aims to classify n observation units into k groups. The observation units in one group have more homogeneous characteristics than the observation units in other groups [7]. The cluster analysis process is to classify the data by using two methods, namely the hierarchical method and the non-hierarchical method. In the hierarchical cluster analysis, it is assumed that at first, each object is a separate cluster, then the two closest things or clusters are combined to form one smaller Cluster [8]. Hierarchical cluster analysis consists of two methods, namely agglomerative and divisive. In the agglomerative method, each object is considered a cluster, then between clusters that are close together and combined into one Cluster. The initial divisive process is all objects comprised in one Cluster then the most different properties are separated and form one other Cluster [8]. The agglomerative method has several algorithms to form clusters, namely single linkage, complete linkage, and average linkage [2]. In this study, the average linkage method was used [8].

According to Hair et al. [4], the concept of similarity is vital in cluster analysis because the principle of cluster analysis is to group objects that have the same characteristics. The distance measure is a measure of similarity. The higher the distance value, the lower the similarity between objects. This research wants to investigate the application of an integrated cluster in SEM with a distance measure, namely the Euclidean distance [9]. Euclidean distance is the most commonly used type of distance measurement because

it is easiest to understand and model. This method is suitable for determining the closest distance between two data. Euclidean distance is the geometric distance between two data objects [8].

2.2 Cluster Validity Index

After conducting cluster analysis, it is necessary to validate the Cluster because the number of groups in the cluster analysis does not yet have a solid basis regarding the number of the best groups. Cluster validation means the procedure for evaluating the results of the cluster analysis quantitatively so that the optimum group is produced. An optimum group is a group that has a dense distance between individuals in the group and is well isolated from other groups [10], [11]. The cluster validity test is carried out to evaluate the results of the quantitative cluster analysis so that the optimum group is produced.

2.2.1 Silhouette

According to Charrad et al. [12], in 1987, Rousseuw and Voicu et al. [13] introduced the Silhouette index with the following equation:

$$S = \frac{\sum_{i=1}^{n} S(i)}{n} \quad (2.1)$$

where,

$$S(i) = \frac{b(i)-a(i)}{max[a(i),b(i)]} \quad (2.2)$$

$a(i) = \frac{\sum d_{ik}}{n_r - 1}$, is the average distance between the ith object and all the observed objects in the same Cluster

$b(i) = min\left(\frac{\sum d_{ik}}{n_s}\right)$ is the average distance of the object with all the observed things contained in the other clusters.

The maximum value of the index is used to determine the number *cluster* optimal.

2.2.2 Krzanowski-Lai

According to Krzanowski-Lai [14], an index that is based on a decrease in the value of the number of squares in a cluster can be defined as follows:

$$DIFF(k) = \left[(k-1)^{1/v} W(k-1) - k^{2/v} W(k)\right] \quad (2.3)$$

then choose one that makes the value below the maximum k

$$KL(k) = \frac{DIFF(k)}{DIFF(k+1)} \quad (2.4)$$

Suppose that it is defined as the number of optimal clusters. If there is an increase in the number of clusters that form up to, the value will decrease drastically. It is expected to be of little value for all values except. It will create the maximum value for optimal.$ggW(k)k < gDIFF(k)kk = gKL(k)k$

2.2.3 Goodman-Kruskal

The Goodman-Kruskal index measures cluster validation internally. The Goodman-Kruskal index finds the concordance and discounting of all possible input parameters. Good clustering is clustering that has many concordant and few concordant [15]. The Goodman-Kruskal index measures the ranking correlation between two sequences A = {a1, a2, .., an} and B = {b1, b2, .., bn} in terms of the number of concordant and discordant pairs in A and B. , aj) and (bi, bj) concordant if they are both ai <aj and bi <bj or ai> aj and bi> bj. Conversely, A and B are discordant if both ai <aj and bi> bj or ai> aj and bi <bj or if, for example, the four pairs of all observed objects are (q, r, s, t) with d (x, y) is the distance between the x and y objects. The four pairs of objects are said to be concordant if they meet the conditions d (q, r) <d (s, t), where q and r are in different groups, and s and t are in the same group. The Goodman-Kruskal index is calculated from the calculation of the value of the concordant and discordant pairs using the formula:

$$GK = \frac{S_c - S_d}{S_c + S_d} \quad (2.5)$$

where, S_c= number of concordant pairs

S_d = number of discordant pairs

Large GK values indicate the optimum group [16].

2.2.4 Dunn

The Dunn validation index denoted by D is calculated by the following formula:

$$D = \min_{1\le i\le n}\left\{\min_{1\le j\le n, i\ne j}\left\{\frac{d(c_i, c_j)}{\max_{1\le k\le n}(d`(c_k))}\right\}\right\} \quad (2.6)$$

where $d(c_i, c_j)$ = distance between ci and cj groups

$d'(c_k)$ = distance in group ck

The most significant value of D is taken as the optimum number of groups [17].

2.2.5 Davies-Bouldin

Davies-Bouldin Index is one of the methods used to measure the validity of clusters in a clustering method; cohesion is defined as the sum of data closeness to the cluster center point of the Cluster being followed. While the separation is based on the distance between the cluster center points to the Cluster. This Davies-Bouldin Index measurement maximizes the inter-cluster space between clusters Ci and Cj and, at the same time, tries to minimize the distance between points in a group. If the inter-cluster length is maximum, the similarities in the characteristics between each Cluster are slight so that the differences between the clusters are more pronounced. If the intra-cluster distance is minimal, it means that each object in the Cluster has a high level of characteristic similarity [18].

$$DBI = \frac{1}{k}\sum_{i=1}^{k} max_{i\ne j}(R_{i,j}) \quad (2.7)$$

where,

$$SSW_i = \frac{1}{m_i}\sum_{j=i}^{m_i} d(x_j, c_j) \quad (2.8)$$

$$SSB_{i,j} = d(c_i, c_j) \quad (2.9)$$

$$R_{i,j} = \frac{SSW_I + SSW_J}{SSB_{i,j}} \quad (2.10)$$

K is the number of clusters used. Based on the above calculations, it can be observed that the smaller the SSW value, the better the clustering results will be obtained. Essentially, the DBI wants a value as small (non-negative ≥ 0) as possible to assess the goodness of the Cluster received. The index is obtained from the average of all cluster indexes, and the value obtained can be used as decision support to assess the number of clusters that are most suitable for use. DBI is also widely used to assist k-means in determining the correct number of clusters to use. Usually, k-means cannot determine the number of clusters used for data clustering [19], [20].

2.3 Linkage

The method that is often used is the agglomerative hierarchy method. The agglomerative hierarchy method algorithms used in group formation include:

2.3.1 Ward Linkage

The variety method aims to obtain clusters that have the most minor possible internal cluster variety. The method commonly used is the Ward method, with the average for each Cluster being calculated. Then, calculate the Euclidean distance between each object and the average value, estimated all the spaces. The two clusters with the increased sum of squares in the smallest Cluster are combined at each stage.

The Ward method forms clusters based on the loss of information due to merging objects into clusters. It is measured using the sum of the squared deviations in the Cluster means for each observation. The sum of squares (SSE) error is used as an objective function. Two objects will be combined if they have the minor objective function among the existing possibilities.

$$SSE = \sum_{j=1}^{p}\left(\sum_{i=1}^{n}\varepsilon_{ij}^2 - \frac{1}{n}\left(\sum_{i=1}^{n}\varepsilon_{ij}^2\right)^2\right) \quad (2.11)$$

where:

ε_{ij} : value for the ith object in *Cluster* j-th

p: number of variables

n: number of respondents in *Cluster* that is formed

2.3.2 Complete Linkage

In the Complete linkage method, the distance between clusters is determined by the farthest distance between two objects in different clusters [8].

$$d_{(ij)k} = max\left(d_{ik}, d_{jk}\right) \quad (2.12)$$

where:

$d_{(ij)k}$: distances between subsample (ij) and cluster k

d_{ik}: distance of subsample i and cluster k

d_{jk}: distance of sub-sample j and Cluster k

2.3.3 Average Linkage Method

In the Average linkage method, the distance between two clusters is considered the average distance between all members in one Cluster and all members of the other clusters. The distance formula can be written in equation (2.6).

$$d_{(ij)k} = \frac{\sum_i \sum_j d_{(ij)}}{N_{ij} N_k} \quad (2.13)$$

where

$d_{(ij)k}$: distances between subsample (ij) and cluster k

d_{ik}: distance of subsample i and cluster k

d_{jk}: distance of sub-sample j and Cluster k

2.4 Structural Equation Modeling (SEM)

SEM is a multivariate analysis, which has the ability and superiority to simultaneously analyze multivariate and multi-relational data, which are relatively complex. SEM is usually used to study the causal relationship between latent variables. SEM has a high degree of flexibility in combining theory and empirical knowledge by modeling observation errors, combining theory and empirical analysis, confirming theory and data (hypothesis testing), and developing approach and data [21]. SEM is a technique used to describe the simultaneous linear relationship between observed variables, which also involves latent variables that cannot be measured directly. SEM analysis combines simultaneous equations, path analysis, regression analysis, and factor analysis [22].

There are two methods in SEM analysis, namely covariance-based and component-based or variant-based methods. Several assumptions that must be fulfilled when using variant-based SEM are multivariate normal distribution, large sample size, and reflective indicators when forming latent variables. SEM analysis based on variance is guided by the fact that theory plays a vital role in constructing the causal relationship of the structural model, and its aim is only to confirm whether the theory-based model differs from the empirical model. Unlike covariance-based SEM, variant-based SEMs such as Partial Least Squares (PLS) and Generalized Structured Component Analysis (GSCA) do not require assumptions. If covariance-based SEM cannot analyze the data, then variant-based SEM can be used [23].

The explanation regarding structural equation modeling will be easier to understand if given an illustration such as Figure 2.1. Figure 2.1. is a structural model, with X1 being the exogenous variable, Y1 being the endogenous mediating variable, and Y2 being the endogenous dependent variable.

(2.12)

2.4.1 SEM with the WarpPLS Approach

The WarpPLS analysis is an extension of the partial least squares (PLS) analysis. PLS is a combination of path and factor analysis and component analysis. PLS is usually referred to as variant-based SEM. If there is a problem with a weak theoretical basis, PLS is the more

appropriate method because it can be used for prediction.

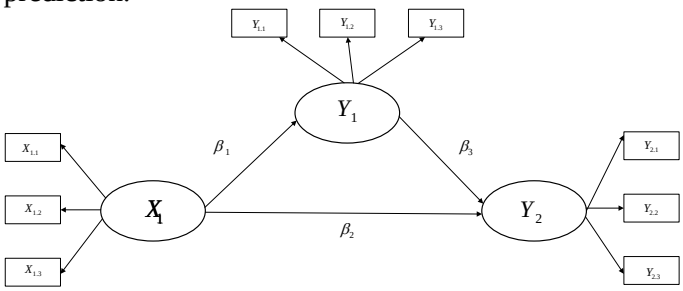

Fig. 1 Structural model illustration

The focus of analysis in PLS shifts from only parameter estimation to validity and accuracy of prediction. It is based on a change in research from estimating model parameters to estimating relevant parameters. There are two characteristics of indicators in PLS, namely reflective indicators and formative indicators. Suppose the structural model to be analyzed is non-recursive, and the latent variables have constructive, thoughtful, or mixed hands. In that case, one of the appropriate methods to be applied is WarpPLS [22]. WarpPLS is a method and package application software program developed by Ned Kock to analyze SEM models based on variants or PLS. WarpPLS software is also equipped with moderating variable analysis with the inconsistent interaction approach.

2.5 Quality of Service

Service quality is a model that describes the condition of customers in forming expectations for service from past experiences, word of mouth promotion, and advertisements by comparing the benefits they expect with what they receive/feel [24]. Meanwhile, service quality is all forms of activities carried out by companies to meet consumer expectations. Service, in this case, is defined as a service or service delivered by the service owner in the form of ease, speed, relationship, ability, and hospitality addressed through attitudes and characteristics in providing services for customer satisfaction. Service quality can be identified by comparing consumers' perceptions of services that are received or obtained with services that are expected or desired for the service attributes of a company. Five indicators can measure service quality, namely: 1) Reliability, 2) Responsiveness, 3) Assurance, 4) Empathy, and 5) Tangibles.

2.6 Environment

According to Munadjat [36], the environment is all objects and conditions, including humans and their actions, which are contained in the space where humans live and affect the life and welfare of humans and other living bodies. In another definition, the environment is defined as a spatial unit with all objects and conditions of living things, including humans, and their behavior and other living things [25]. Meanwhile, according to Robbins and Stephen [26], the environment is institutions or outside forces that can affect organizational performance; the environment is formulated into two: the general environment and the unique environment. The general environment is anything outside the organization that has the potential to influence the organization. This environment is in the form of social and technological conditions. In contrast, the unique environment is the part of the environment directly related to achieving an organization's goals. Two indicators can measure the environment, namely: 1) Physical Environment, and 2) Non-Physical Environment.

2.7 Fashion

Currently, fashion is related to clothing or clothes, when in fact, what is said to be fashion is everything that is trending in society. It includes clothing, appetite, entertainment, consumer goods, and so on. According to Hass [27] in his book, Sociology, "fashion is a great though brief enthusiasm among the relatively large number of people for a particular innovation." So actually fashion can include anything that is followed by many people and becomes a trend. Fashion is also related to novelty or novelty elements. Therefore fashion tends to be short-lived and not eternal. And because what tends to move and change every time is clothing, fashion is often associated with clothing. In contrast, as long as something new about an artifact involves the fun of many people, it can become a fashion [27]. Two indicators can measure the environment: 1) Activities, 2) Interests, and 3) Opinions.

2.7 *Willingness to Pay*

*Willingness to pay*ments the willingness of the credit applied to pay the payment burden according to a predetermined amount of credit. Willingness to pay or willingness to pay is closely related to variables that affect the ability of lenders to force payments through legal compensation; it is an integral part of the overall effectiveness analysis process and creditor friendliness [28], [29]. Willingness to pay is a value where someone is willing to pay, sacrifice or exchange something to obtain goods or services. According to Permadi et al. [30], five indicators can measure willingness to pay, namely: 1) Consultation before making payments; 2) Documents required to pay; 3) Information regarding the method and place of payment; 4) Information regarding the payment deadline; 5) Allocate payment funds.

2.8 Compliant Paying Conduct

According to Budiono [31] and Boccia et al. [32], behavior is a human reaction or action caused by an impulse seen from values, habits, driving forces, motives, and strength of detention endogenous or

someone's reaction that appears. It is due to the experience of the stimulation process and learning from the environment. According to Sahu [33], obedience is an attitude of being willing to do whatever is based on self-awareness or coercion, which causes behavior according to what is not expected. Obedient behavior is the interaction of individual, organizational, and group behavior [34]. Compliant paying behavior is defined as someone's action caused by self-awareness or compulsion to pay their obligations. According to Bambang and Wide [35], three indicators can measure customer compliance, namely: 1) Timeliness; 2) Accuracy of data; 3) Sanctions.

3. Methods and Materials

This study uses primary data; the variables used are service quality, environment, fashions, willingness to pay, and compliance with paying behavior at Bank X. The data consists of three exogenous variables: service quality, environment, fashion, and two endogenous variables: willingness to pay and compliant pay behavior. Data obtained through a questionnaire with a Likert scale. Measurement of variables in primary data using the average score of each item. The sampling technique used was purposive sampling. Purposive sampling is a sampling technique based on specific characteristics or conditions that are the same as the characteristics of the population. The sample used is 100 Bank X customers.

Data analysis was carried out quantitatively to explain each of the variables studied; a descriptive analysis was carried out first, then carried out by Structural Equation Modeling (SEM) analysis based on Partial Least Square (PLS). This study was used as an analysis tool. According to Solimun et al. [22], using the WarpPLS program will obtain a PLS (Partial Least Square) analysis model for the following reasons: (1) The analysis model is tiered, and the structural equation model meets the recursive model. (2) Measurement of latent variables, namely any variables that cannot be measured directly.

The following is a picture of the research hypothesis model obtained based on the results of previous research.

Based on the research hypothesis model in Figure 1, the research hypothesis can be formulated as follows:

H1: Service Quality (X1) has a significant effect on willingness to pay

H2: Environment (X2) has a substantial impact on *Willingness To Pay*

H3: Fashions (X3) has a substantial impact on *Willingness To Pay*

H4: Quality of Service (X1) has a substantial impact on Compliant Paying Conduct

H5: Environment (X2) has a substantial impact on Compliant Paying Conduct

H6: *Fashions* (X3) has a substantial impact on Compliant Paying Conduct

H7: *Willingness To Pay* (Y1) has a substantial impact on Compliant Paying Conduct

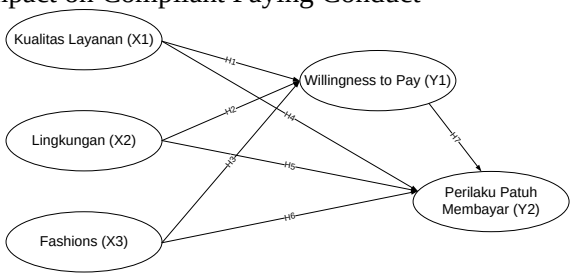

Fig. 1 The research hypothesis model

4. Results and Discussion

4.1 Cluster Analysis

This study uses 8 cluster validity indices. This study indicates that the number of Clusters 1 and 2 for all indexes has the same number; namely, for cluster 1, there are 42 members, and Cluster 2 has 58 members of the cluster validity index. The average results obtained can be seen in Table 1.

It can be seen from Table 1, the Cluster mean for the index Silhouette, Krzanowski-Lai, Dunn, and Davies-Bouldin had the same mean results for each linkage. It means that the Silhouette, Krzanowski-Lai, Dunn, and Davies-Bouldin indexes have the same cluster members for each link.

Thus, in conducting SEM analysis of the Warp-PLS approach, the researcher uses the Silhouette index with an average and complete linkage representing the Krzanowski-Lai, Dunn, and Davies-Bouldin index for each linkage because they have the same members of each Cluster.

4.2 Integrated Cluster Index Model *Silhouette* and Ward Linkage

Based on the results of cluster analysis with the Silhouette index and ward linkage, it was found that the number of clusters was 2 clusters, with Cluster 1 as many as 36 customers and Cluster 2 as many as 64 customers. Next, a dummy will be formed from the resulting clusters. The number of clusters formed is 2 clusters, so there is one dummy. The researcher determines customers in Cluster 1 as Dummy 1 and customers in Cluster 2 as Dummy 0.

Table 1 Average Cluster Members for Index and Linkage respectively

Method	Average									
	X1		X2		X3		Y1		Y2	
	C1	C2	C1	C2	C1	C2	C1	C2	C1	C2
Silhouette Index										
Ward	4.022	3.200	4.049	3.191	4.060	3.224	4.072	3.145	4.000	3.201

Linkage										
Average Linkage	3.943	3.172	4.048	3.103	3.929	3.233	3.967	3.126	4.004	3.115
Complete Linkage	3.754	3.217	4.010	2.948	3.718	3.316	3.819	3.110	3.843	3.104
Krzanowski-Lai Index										
Ward Linkage	4.022	3.200	4.049	3.191	4.060	3.224	4.072	3.145	4.000	3.201
Average Linkage	3.943	3.172	4.048	3.103	3.929	3.233	3.967	3.126	4.004	3.115
Complete Linkage	3.754	3.217	4.010	2.948	3.718	3.316	3.819	3.110	3.843	3.104
Dunn Index										
Ward Linkage	4.022	3.200	4.049	3.191	4.060	3.224	4.072	3.145	4.000	3.201
Average Linkage	3.943	3.172	4.048	3.103	3.929	3.233	3.967	3.126	4.004	3.115
Complete Linkage	3.754	3.217	4.010	2.948	3.718	3.316	3.819	3.110	3.843	3.104
Davies-Bouldin Index										
Ward Linkage	4.022	3.200	4.049	3.191	4.060	3.224	4.072	3.145	4.000	3.201
Average Linkage	3.943	3.172	4.048	3.103	3.929	3.233	3.967	3.126	4.004	3.115
Complete Linkage	3.754	3.217	4.010	2.948	3.718	3.316	3.819	3.110	3.843	3.104

Model feasibility test or Goodness of Fit testing the fit/suitability of the model with the research data. The goodness of fit in question is an index or measure of the integrity of the relationship between latent variables (inner model) related to its assumptions. In this study, the criteria in determining the goodness/ feasibility of the model for an integrated cluster with the SEM Warp-PLS approach can be seen in Table 2.

Table 2 summarizes the results obtained in the analysis and the recommended values for measuring the feasibility of the model. Based on the results of the feasibility test of the model as a whole, not all of the criteria reached the expected value limit or did not meet the recommended goodness of fit indices critical limit, so the results of this modeling could not be accepted or worthy of analysis. Several criteria were rejected, including Average block VIF, Average full collinearity VIF, Sympson's paradox ratio, and R-squared contribution ratio. It can be stated that this test results in a poor conformation of the variables as well as the causal relationship between variables. Thus, the overall model test shows unfavorable results or following expectations, meaning that the empirical data (field data) does not support the theoretical model developed.

Table 2 Model Feasibility test results for integrated clusters with SEM Approach Warp-PLS Silhouette Index and Ward Linkage

No.	Model Fit / Quality Index	Score	Criteria	Information
1	Average path coefficient	APC = 0.474P <0.001	P <0.05	Significant
2	Average R-squared	ARS = 0.708 P <0.001	P <0.05	Significant
3	Average adjusted R-squared	AARS = 0.720P <0.001	P <0.05	Significant
4	Average block VIF	AVIF = 83,504	*acceptable* if AVIF ≤ 5 ideal if AVIF ≤ 3.3	Rejected
5	Average full collinearity VIF	AFVIF = 196,167	*acceptable if AFVIF ≤ 5* *ideal if AFVIF ≤ 3.3*	Rejected
6	Tenenhaus GoF	GoF = 0.748	*small if GoF ≥ 0.1* *medium if GoF ≥ 0.25* *large if GoF ≥ 0.36*	Big
7	Sympson's paradox ratio	SPR = 0.563	*acceptable if the SPR ≥ 0.7* *ideal if SPR = 1*	Rejected
8	R-squared contribution ratio	RSCR = 0.368	*acceptable if RSCR ≥ 0.9* *ideal RSCR = 1*	Rejected
9	Statistical suppression ratio	SSR = 0.875	*acceptable if SSR ≥ 0.7*	Acceptable
10	Nonlinear bivariate causality direction ratio	NLBCDR = 1,000	*acceptable if NLBCDR ≥ 0.7*	Acceptable

4.3 Silhouette Index and Average Linkage Integrated Cluster Model

Based on the results of cluster analysis with the silhouette index and average linkage, it was found that the number of clusters was 2 clusters, with cluster 1 as many as 42 customers and cluster 2 as many as 58 customers. Next, a dummy will be formed from the

resulting clusters. The number of clusters formed is 2 clusters, so there is one dummy. The researcher determines customers in Cluster 1 as Dummy 1 and customers in Cluster 2 as Dummy 0.

Model feasibility test or Goodness of Fit testing the fit/suitability of the model with the research data. The goodness of fit in question is an index or measure of the integrity of the relationship between latent variables (inner model) related to its assumptions. In this study, the criteria for determining the goodness/feasibility of the model for an integrated cluster with the SEM Warp-PLS approach can be seen in Table 3.

Table 3 Model Feasibility test results for an integrated cluster with SEM Approach Warp-PLS Silhouette Index and Average Linkage

No.	Model Fit / Quality Index	Score	Criteria	Information
1	Average path coefficient	APC = 0.607P <0.001	P <0.05	Significant
2	Average R-squared	ARS = 0.627 P <0.001	P <0.05	Significant
3	Average adjusted R-squared	AARS = 0.885P <0.001	P <0.05	Significant
4	Average block VIF	AVIF = 123,659	*acceptable* if AVIF ≤ 5 ideal if AVIF ≤ 3,3	Rejected
5	Average full collinearity VIF	AFVIF = 129,380	*acceptable if AFVIF ≤ 5* *ideal if AFVIF ≤ 3,3*	Rejected
6	Tenenhaus GoF	GoF = 0.133	*small if GoF ≥ 0.1* *medium if GoF ≥ 0.25* *large if GoF ≥ 0.36*	Small
7	Sympson's paradox ratio	SPR = 0.625	*acceptable if the SPR ≥ 0.7* *ideal if SPR = 1*	Rejected
8	R-squared contribution ratio	RSCR = 0.264	*acceptable if RSCR ≥ 0.9* *ideal RSCR = 1*	Rejected
9	Statistical suppression ratio	SSR = 0.813	*acceptable if SSR ≥ 0.7*	Acceptable
10	Nonlinear bivariate causality direction ratio	NLBCDR = 1,000	*acceptable if NLBCDR ≥ 0.7*	Acceptable

Table 3 summarizes the results obtained in the analysis and the recommended values for measuring the feasibility of the model. Based on the results of the feasibility test of the model as a whole, not all of the criteria reached the expected value limit or did not meet the recommended goodness of fit indices critical limit, so the results of this modeling could not be accepted or worthy of analysis. Several criteria were rejected, including Average block VIF, Average full collinearity VIF, Sympson's paradox ratio, and R-squared contribution ratio. It can be stated that this test results in a poor conformation of the variables and the causal relationship between variables. So, the overall model test shows unfavorable results or following expectations.

4.4 Silhouette Index and Complete Linkage Integrated Cluster Model

Based on the results of cluster analysis with the Silhouette index and complete linkage, it was found that the number of clusters was 2 clusters, with Cluster 1 as many as 36 customers and Cluster 2 as many as 64 customers. Next, a dummy will be formed from the resulting clusters. The number of clusters formed is 2 clusters, so there is one dummy. The researcher determines customers in Cluster 1 as Dummy 1 and customers in Cluster 2 as Dummy 0.

Model feasibility test or Goodness of Fit testing the fit/suitability of the model with the research data. The goodness of fit in question is an index or measure of the integrity of the relationship between latent variables (inner model) related to its assumptions. In this study, the criteria in determining the goodness/feasibility of the model for an integrated cluster with the SEM Warp-PLS approach can be seen in Table 4.

Table 4 summarizes the results obtained in the analysis and the recommended values for measuring the feasibility of the model. Based on the results of testing the feasibility of the model as a whole, all the criteria reached the expected value limit or met the recommended Goodness of fit indices critical limit. The modeling results can be accepted or worthy of analysis. Several criteria were rejected, including Average block VIF and Average full collinearity VIF. It can be stated that this test results in good confirmation of the variables as well as the causality relationship between variables. So, the overall model test shows good results or following expectations, meaning that the empirical data (field data) already supports the

Table 4 Model Feasibility Test Results for Integrated Cluster with SEM Approach Warp-PLS Silhouette Index and Complete Linkage

No.	Model Fit / Quality Index	Score	Criteria	Information
1	Average path coefficient	APC = 0.165P = 0.022	P <0.05	Significant
2	Average R-squared	ARS = 0.771 P <0.001	P <0.05	Significant
3	Average adjusted R-squared	AARS = 0.749P <0.001	P <0.05	Significant
4	Average block VIF	AVIF = 12.021	*acceptable* if AVIF ≤ 5	Rejected

No.	Model Fit / Quality Index	Score	Criteria	Information
			ideal if AVIF ≤ 3.3	
5	Average full collinearity VIF	AFVIF = 96,879	*acceptable if AFVIF ≤ 5* *ideal if AFVIF ≤ 3.3*	Rejected
6	Tenenhaus GoF	GoF = 0.778	*small if GoF ≥ 0.1* *medium if GoF ≥ 0.25* *large if GoF ≥ 0.36*	Big
7	Sympson's paradox ratio	SPR = 0.750	*acceptable if the SPR ≥ 0.7* *ideal if SPR = 1*	Ideal
8	R-squared contribution ratio	RSCR = 0.923	*acceptable if RSCR ≥ 0.9* *ideal RSCR = 1*	Acceptable
9	Statistical suppression ratio	SSR = 1,000	*acceptable if SSR ≥ 0.7*	Acceptable
10	Nonlinear bivariate causality direction ratio	NLBCDR = 1,000	*acceptable if NLBCDR ≥ 0.7*	Acceptable

the theoretical model being developed.

4.5 Comparison of the R2 SEM Value of the Warp-PLS approach on the Integrated Cluster with Various Indices and Linkages

In this study, the criteria for determining the best model for an integrated cluster with SEM with the Warp-PLS approach at various indexes and linkages can be seen in Table 5.

Based on Table 5, it can be seen that the integrated cluster model of the SEM Warp-PLS approach using the Silhouette, Krzanowski-Lai, Dunn, and Davies-Bouldin indexes with complete linkage has an R2 value of 0.749 which means the variable service quality, environment, fashions, and willingness to Pay simultaneously affects the respectful behavior of paying by 74.9%. In comparison, other variables influence the remaining 26.1%. The integrated cluster model of the SEM Warp-PLS approach using the Silhouette, Krzanowski-Lai, Dunn, and Davies-Bouldin indexes with ward linkage has an R2 value 0.720, which means that the variables of service quality, environment, fashions, and willingness to pay simultaneously affect compliance behavior pay 72.0%. In comparison, other variables influence the remaining 28.0%. The integrated cluster model of the SEM Warp-PLS approach using the Silhouette, Krzanowski-Lai, Dunn, and Davies-Bouldin indexes with average linkage having an R2 value of 0.627 which means that the service quality, environment, fashions, and willingness to pay variables simultaneously affect compliance behavior pay 62.7%, while other variables influence the remaining 37.3%. Based on the R2 value, it can be concluded that the Silhouette, Krzanowski-Lai, Dunn, and Davies-Bouldin indexes with complete linkage are the best indexes with a structural model that can be seen in the following equation: and willingness

to pay simultaneously affect compliance with paying behavior by 62.7%. In comparison, other variables influence the remaining 37.3%. Based on the R2 value, it can be concluded that the Silhouette, Krzanowski-Lai, Dunn, and Davies-Bouldin indexes with complete linkage are the best indexes with a structural model that can be seen in the following equation: and willingness to pay simultaneously affect compliance with paying behavior by 62.7%. In comparison, other variables influence the remaining 37.3%.

Table 5 Comparison of R2 Value

Index	R2 value
Silhouette	
Ward	0.720
Average	0.627
Complete	0.749
Krzanowski-Lai	
Ward	0.720
Average	0.627
Complete	0.749
Dunn	
Ward	0.720
Average	0.627
Complete	0.749
Davies-Bouldin	
Ward	0.720
Average	0.627
Complete	0.749

Based on the R2 value, it can be concluded that the Silhouette, Krzanowski-Lai, Dunn, and Davies-Bouldin indexes with complete linkage are the best indexes with a structural model that can be seen in the following equation:

$$Y_1 = \beta_1 D_1 + \beta_2 X_1 + \beta_3 X_2 + \beta_4 X_3 + \beta_5 D_2 X_1 + \beta_6 D_3 X_2 + \beta_7 D_4 X_3$$

$$Y_1 = 0,142 D_1 + 0,354 X_1 + 0,001 X_2 + 0,001 X_3 + 0,363 D_2 X_1 + 0,241 D_3 X_2 + 0,279 D_4 X_3$$

Cluster 1 (D = 1):

$$Y_1 = 0,142 + 0,717 X_1 + 0,242 X_2 + 0,280 X_3 \quad (4.1)$$

Cluster 2 (D = 0):

$$Y_1 = 0,354 X_1 + 0,001 X_2 + 0,001 X_3 \quad (4.2)$$

Based on Equations 4.1 and 4.2, it can be concluded that service quality (X1), fashions (X3), and environment (X2) in cluster 1 have a more significant influence than cluster 2. In Cluster 1, each increase is one unit of environmental quality (X1) it will increase the customer's willingness to pay (Y1) by 0.717 units. Every increase of one environmental company (X2) for a customer in Cluster 1 will increase the customer's willingness to pay (Y1) by 0.242 units. Also, each increase of one customer's fashions (X3) unit in cluster 1 will increase the customer's willingness to pay (Y1) by 0.280 units.

In Cluster 2, each increase of one environmental quality unit (X1) will increase the customer's willingness to pay (Y1) by 0.354 units. Every increase of one environmental team (X2) of a customer in Cluster 2 will increase the customer's willingness to pay (Y1) by 0.001 units. Also, each increase of one customer's fashion unit (X3) in Cluster 2 will increase the customer's willingness to pay (Y1) by 0.001 units.

$$Y_2 = \beta_1 D_1 + \beta_2 X_1 + \beta_3 X_2 + \beta_4 X_3 + \beta_5 Y_1 + \beta_6 D_2 X_1 + \beta_7 D_3 X_2 + \beta_8 D_4 X_3 + \beta_9 D_5 Y_1$$

$$Y_2 = 0,284 D_1 + 0,001 X_1 + 0,042 X_2 + 0,167 X_3 + 0,326 Y_1 + 0,417 D_2 X_1 + 0,195 D_3 X_2 + 0,001 D_4 X_3 + 0,469 D_5 Y_1$$

Cluster 1 (D = 1):

$$Y_2 = 0,284 + 0,418 X_1 + 0,237 X_2 + 0,168 X_3 + 0,795 Y_1 \quad (4.3)$$

Cluster 2 (D = 0):

$$Y_2 = 0,001 X_1 + 0,042 X_2 + 0,167 X_3 + 0,326 Y_1 \quad (4.4)$$

Based on Equations 4.3 and 4.4, it can be concluded that service quality (X1), environment (X2), fashions (X3), and willingness to pay (Y1) in Cluster 1 have a more significant influence than Cluster 2. In Cluster 1, each increase in one unit of environmental quality (X1) will increase customers' compliance to pay behavior (Y2) by 0.418 companies. Every increase of one environmental unit (X2) of customers in Cluster 1 will increase the customer compliance behavior (Y2) of 0.237 units. Every increase of one unit of customer fashions (X3) in Cluster 1, will increase the customer compliance behavior (Y2) by 0.168 units. Also, every increase of one unit of willingness to pay (Y1) of customers in Cluster 1, will increase the compliance behavior of customers (Y2) by 0.795 units.

In Cluster 2, each increase of one environmental quality unit (X1) will increase customers' compliance to pay behavior (Y2) by 0.001 units. Every increase of one environmental unit (X2) of customers in Cluster 2 will increase the customer compliance behavior (Y2) of 0.042 units. Every increase of one unit of customer fashions (X3) in Cluster 2, will increase the customer compliance behavior (Y2) by 0.167 units. Also, every increase of one unit of willingness to pay (Y1) of customers in cluster 1, will increase customer compliance to pay behavior (Y2) by 0.326 units.

5. Conclusion

Based on the analysis results, the conclusion that can be given is applying an integrated cluster in SEM with the Warp-PLS approach with various cluster validity index methods resulting in many clusters and the same cluster members causing the same dummy variables. The R2 value in the integrated Cluster with the Silhouette, Krzanowski-Lai, Dunn, and Davies-Bouldin indexes, the complete linkage method is better than the ward and average linkage methods. Variable Service quality (X1), environment (X2), fashions (X3), and willingness to pay (Y1) in Cluster 1 have a more significant influence than Cluster 2.

Suggestions that can be given are based on the results of the integrated Cluster on SEM with the Warp-PLS approach, namely for further research to compare the effect of using distance, index, and other linkages on the discriminant integrated cluster analysis which results in a high R2 value.

References

[1] SOLIMUN M. S. *Structural Equation Modeling (SEM) with LISREL and AMOS*. Malang: Faculty of Mathematics and Natural Sciences, Brawijaya University, 2002.

[2] SUPRANTO J. *Multivariate analysis of meaning and interpretation*. Jakarta: PT Rineka Cipta, 2004.

[3] SOLIMUN M. S. *Multivariate analysis of structural modeling*. Malang: CV Image of Malang, 2010.

[4] HAIR J.F., ANDERSON R.E., TATHAM R.L. et al. *Multivariate data analysis*. 5th edition. Jakarta: Gramedia Main Library, 2006.

[5] LIU Q., ZHANG R., HU R. et al. An improved path-based clustering algorithm. *Knowledge-Based Systems*, 2018, 163. doi: 10.1016/j.knosys.2018.08.012.

[6] JAIN A.K., AMBASSADOR R.C. *Algorithms for data clustering*. New Jersey: Prentice-Hall, 1988.

[7] MATJIK A., SUMERTAJAYA I. *Perancangan dan Percobaan dengan Aplikasi SAS dan Minitab. (Design and Experimentation with SAS and Minitab Applications)*. 2nd ed. Bogor: IPB Press, 2002.

[8] JOHNSON R.A., WICHERN D.W. *Applied multivariate analysis*. 3rd ed. New Jersey: Prentice-Hall, 1992.

[9] FU W., PERRY P. O. Estimating the number of clusters using cross-validation. *Journal of Computational and Graphical Statistics*, 2020, 29(1): 162-173.

[10] JAIN A. K., DUBER R. C. *Algorithms for clustering data*. New Jersey: Prentice-Hall, 1988.

[11] PATIBANDLA R. L., VEERANJANEYULU, N. Performance analysis of partition and evolutionary clustering methods on various cluster validation criteria. *Arabian Journal for Science and Engineering*, 2018, 43(8):379-439.

[12] CHARRAD M., GHAZZALI N., BOITEAU V. et al. Package 'nbclust'. *Journal of Statistical Software*, 2014, 61 (6): 1-36.

[13] VOICU A., DUTEANU N., VOICU M., et al. The rock and cluster R packages applied to drug candidate selection. *Journal of Cheminformatics*, 2020, 12(1): 1-8.

[14] KRAZANOWSKI W. J. & LAI Y.T. A criterion for determining the number of groups in a data set using a sum of squares clustering. *Biometrics*, 1988, 44: 23–34.

[15] GOODMAN L. A., & KRUSKAL W. H. Measures of association for cross classifications. *Journal of the American Statistical Association*, 1954, 49: 732–769.

[16] BOLSHAKOVA N. & AZUAJE F. Cluster validation techniques for genome expression data. *Signal Processing*, 2003, 83 (4): 825-833.
[17] AZUAJE F., BOLSHAKOVA N. Clustering genomic expression data: Design and evaluation principles. In: BERRAR D.P., DUBITZKY W., GRANZOW M. (Eds.) *A practical approach to microarray data analysis*. Boston, MA: Springer, 2003. DOI: 10.1007/0-306-47815-3_13.
[18] WANI M.A. & RIYAZ R. A novel point density-based validity index for clustering gene expression datasets. *International Journal of Data Mining and Bioinformatics*, 2017, 17 (1): 66-84.
[19] BATES A. & KALITA J. Counting clusters in Twitter posts. In: *Proceedings of the Second International Conference on Information and Communication Technology for Competitive Strategies (ICTCS '16)*. New York: Association for Computing Machinery, 2016, Article 85: 1–9. DOI: 10.1145/2905055.2905295.
[20] RUZ G. A., HENRÍQUEZ P. A., & MASCAREÑO A. Sentiment analysis of Twitter data during critical events through Bayesian networks classifiers. *Future Generation Computer Systems*, 2020, 106: 92-104.
[21] WANG J. & WANG X. *Structural equation modeling: Applications using Mplus*. New York: John Wiley & Sons, 2019.
[22] SOLIMUN F., NURJANNAH N. *Multivariate Statistical Method: Structural Equation Modeling Based on WarpPLS, 0. 90*. Malang: UB Press, 2017.
[23] GHOZALI I. *Structural equation modeling: An alternative method with partial least squares (PLS)*. Semarang: Diponegoro University Publishing Agency, 2008.
[24] RAMYA, N., KOWSALYA, A., & DHARANIPRIYA, K. Service quality and its dimensions. *EPRA International Journal of Research & Development*. 2019, 4: 38-41.
[25] SOERIAATMADJA R.E. *Environmental Science*. Bandung: ITB Publisher, 1997.
[26] ROBBINS S.P. *Organizational behavior: controversy, concept, application*. Jakarta: Prehalinda, 2002.
[27] HASS J. K. *Economic sociology: An introduction*. Abingdon: Routledge, 2020.
[28] GOLIN J. & DELHAISE P. *The bank credit analysis handbook: a guide for analysts, bankers, and investors*. New York: John Wiley & Sons, 2013.
[29] PALACÍN-SÁNCHEZ M. J., CANTO-CUEVAS F. J., & DI-PIETRO F. Trade credit versus bank credit: a simultaneous analysis in European SMEs. *Small Business Economics*, 2019, 53(4): 1079-1096.
[30] PERMADI T., NASIR A. & ANISMA Y. Study of willingness to pay taxes on individual taxpayers who do independent work (the case at KPP Pratama Tampan Pekanbaru). *Journal of Economics*, 2013, 21 (02):43-52.
[31] BUDIONO D. *The behavior of corporate taxpayers in fulfilling tax obligations: Humanistic theory perspective.* Jakarta: FEBSOS, 2016.
[32] BOCCIA F., MALGERI M. R. & COVINO D. Consumer behavior and corporate social responsibility: An evaluation by a choice experiment. *Corporate Social Responsibility and Environmental Management*, 2019, 26(1), 97-105.
[33] SAHU R. A gap analysis of enforcing FCPA-compliant codes of conduct in low-and middle-income countries. *Business Law Review*, 2020, 41(6):56-64.
[34] SIAT C.C., TOLY A.A. Factors affecting taxpayer compliance in fulfilling tax obligations in Surabaya. *Tax & Accounting Review*, 2013, 1(1): 41.
[35] SUBROTO B. *The influence of attitudes, subjective norms, perceived behavior control, and sunset policy on taxpayer compliance with intention as an intervening variable*. Melville, NY: JMP Press, 2010
[36] MUNADJAT D. *Environmental law (book v: sectoral): Indonesian environmental law (in the national & international system).* Bandung: Binacipta, 1984.

参考文:

[1] SOLIMUN M. S. 具有 LISREL 和 AMOS 的結構方程建模（SEM）。 瑪瑯：布拉維再也大學數學與自然科學學院，2002。

[2] SUPRANTO J.含義和解釋的多元分析。雅加達：PT Rineka Cipta，2004。

[3] SOLIMUN M. S. 結構建模的多元分析。瑪瑯：瑪瑯的簡歷圖片，2010。

[4] HAIR J.F.，ANDERSON R.E.，TATHAM R.L.等。多元數據分析。第 5 版。雅加達：Gramedia 主圖書館，2006 年。

[5] LIU Q., ZHANG R., HU R. 等。一種改進的基於路徑的聚類算法。知識庫系統，2018，163. doi：10.1016 / j.knosys.2018.08.012。

[6] JAIN A.K., AMBASSADOR R.C. 數據聚類算法。新澤西州：普倫蒂斯·霍爾（Prentice-Hall），1988。

[7] MATJIK A.，SUMERTAJAYA I. Perancangan dan 和 Percobaan dengan Aplikasi SAS 以及 Minitab。（使用 SAS 和 Minitab 應用程序進行設計和試驗）。第二版。茂物：IPB 出版社，2002 年。

[8] JOHNSON R.A.，WICHERN D.W.應用多元分析。第三版。新澤西州：普倫蒂斯·霍爾（Prentice Hall），1992 年。

[9] FU W.，PERRY P. O.使用交叉驗證估計聚類數。計算與圖形統計學報，2020，29（1）：162-173。
[10] JAIN A. K.，DUBER R. C. 數據聚類算法。新澤西州：普倫蒂斯·霍爾（Prentice-Hall），1988。
[11] PATIBANDLA R. L.，VEERANJANEYULU，N。基於各種聚類驗證標準的分區和進化聚類方法的性能分析。阿拉伯科學與工程學報，2018，43（8）：379-439。
[12] CHARRAD M.，GHAZZALI N.，BOITEAU V.等。包'nbclust'。統計軟件學報，2014，61（6）：1-36。
[13] VOICU A.，DUTEANU N.，VOICU M.等。 Rock and Cluster R 軟件包適用於候選藥物的選擇。化學信息學報，2020，12（1）：1-8。
[14] KRAZANOWSKI W. J.＆LAI Y.T. 使用平方和聚類確定數據集中的組數的標準。生物識別，1988，44：23-34
[15] GOODMAN L. A. 和 KRUSKAL W. H. 交叉分類的關聯度量。美國統計協會雜誌，1954，49：732–769。
[16] BOLSHAKOVA N. 和 AZUAJE F. 基因組表達數據的聚類驗證技術。信號處理，2003，83（4）：825-833。
[17] AZUAJE F.，BOLSHAKOVA N. 聚類基因組表達數據：設計和評估原則。在：BERRAR D.P.，DUBITZKY W.，GRANZOW M.（編輯）的一種實用方法進行微陣列數據分析。馬薩諸塞州波士頓：斯普林格，2003 年。DOI：10.1007 / 0-306-47815-3_13。
[18] WANI M.A. 和 RIYAZ R.一種新穎的基於點密度的聚類基因表達數據集有效性指數。國際數據挖掘與生物信息學雜誌，2017，17（1）：66-84。
[19] BATES A.和 KALITA J.對 Twitter 帖子中的簇進行計數。於：第二屆信息和通信技術國際競爭戰略國際會議論文集（ICTCS '16）。紐約：計算機協會，2016 年，第 85 條：1–9。 DOI：10.1145 / 2905055.2905295。
[20] RUZ G. A.，HENRÍQUEZP. A. 和 MASCAREÑOA. 在關鍵事件期間通過貝葉斯網絡分類器對 Twitter 數據進行情感分析。下一代計算機系統，2020，106：92-104。
[21] WANG J. ＆WANG X. 結構方程建模：使用 Mplus 的應用。紐約：約翰·威利父子（John Wiley＆Sons），2019 年。
[22] SOLIMUN F. 和 NURJANNAH N.多元統計方法：基於 WarpPLS 的結構方程模型，0。90. Malang：UB 出版社，2017 年。
[23] GHOZALI I. 結構方程建模：偏最小二乘（PLS）的另一種方法。三寶壟：Diponegoro 大學出版社，2008.
[24] RAMYA, N., KOWSALYA, A., 和 DHARANIPRIYA K。服務質量及其規模。 EPRA 國際研究與發展雜誌。 2019，4：38-41。
[25] SOERIAATMADJA R.E. 環境科學。萬隆：ITB 出版商，1997 年。
[26] ROBBINS S.P. 組織行為：爭議，概念，應用。雅加達：Prehalinda，2002。
[27] HASS J. K. 經濟社會學：導論。阿賓登：魯特利奇（Routledge），2020。
[28] GOLIN J. 和 DELHAISE P. 銀行信貸分析手冊：面向分析師，銀行家和投資者的指南。紐約：約翰·威利父子（John Wiley＆Sons），2013。
[29] PALACÍN-SÁNCHEZM. J.，CANTO-CUEVAS F. J. 和 DI-PIETRO F. 貿易信貸與銀行信貸：歐洲中小型企業的同步分析。小型企業經濟學，2019，53（4）：1079-1096。
[30] PERMADI T.，NASIR A. 和 ANISMA Y. 對從事獨立工作的個人納稅人是否願意納稅的研究（KPP Pratama Tampan Pekanbaru 的情況）。經濟學雜誌，2013，21（02）：43-52。
[31] BUDIONO D. 公司納稅人在履行稅收義務方面的行為：人文主義理論視角。雅加達：FEBSOS，2016.
[32] BOCCIA F.，MALGERI M. R. 和 COVINO D. 消費者行為與企業社會責任：通過選擇實驗進行的評估。企業社會責任與環境管理，2019，26（1），97-105。
[33] SAHU R.在中低收入國家中執行 FCPA 行為守則的差距分析。 商法評論，2020，41（6）：56-64。
[34] SIAT C.C.，TOLY A.A. 影響泗水履行納稅義務的納稅人合規性的因素。 稅務與會計評論，2013，1（1）：41。
[35] SUBROTO B. .態度，主觀規範，感知的行為控制和日落政策對納稅人是否遵守意向作為乾預變量的影響。 紐約州梅爾維爾：JMP 出版社，2010.
[36] MUNADJAT D. 環境法（第五卷：部門法）：印度尼西亞環境法（在國內和國際體系中）。 萬隆（Bandung）：比那西普塔（Binacipta），1984。