Journal of Theoretical and Applied Information Technology
31st January 2023. Vol.101. No 2

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

CREDIT CUSTOMER SEGMENTATION WITH HIERARCHICAL CLUSTERING AT VARIOUS DISTANCES

HENIDA RATNA AYU PUTRI[1*], ADJI ACHMAD RINALDO FERNANDES[2], ATIEK IRIANY[3], NURJANNAH[4], SOLIMUN[5]

[1,2,3,4,5]Department of Statistics, Faculty of Mathematics and Natural Sciences, Brawijaya University
Corresponding author e-mail: henida_ratna@student.ub.ac.id [1*)]

ABSTRACT

Cluster analysis is a multivariate technique that aims to group objects based on similar characteristics. The purpose of this study was to see the grouping and comparison of the results of cluster analysis using the ward linkage method with various distances (Euclidean, Manhattan and Mahalanobis) with an assessment of the 5C variable (Character, Capacity, Capital, Collateral, Condition of Economy) on Bank X mortgage customers. In this study, we will compare the results of grouping using the three methods based on the ratio of standard deviations between clusters and within clusters. The sampling technique used nonprobability sampling method with purposive sampling basis. The sample size in this study was 100 customers. The standard deviation ratio between clusters and within clusters shows that the results of the Ward method grouping with Mahalanobis distance are better than the results of grouping using other methods. The results of cluster analysis using the Ward method with Mahalanobis distance are cluster one consisting of 26 customers with compliance as the highest aspect, cluster two consisting of 30 customers with condition as the highest aspect, while cluster three consists of 44 customers with capital as the highest aspect. The originality of this study is to compare the distances in hierarchical cluster analysis, especially with ward linkage in segmentation of credit customer.
Keywords: Cluster analysis, Euclidean distance, Manhattan distance, Mahalanobis distance, Ward's Method

1. INTRODUCTION

Narimawati [1] said that cluster analysis is a multivariate technique that aims to classify objects into groups that differ from one group to another. Objects that have been classified in one cluster are objects that have the same relative distance proximity to other objects. In cluster analysis, there are two methods that can be used to form a cluster, namely the hierarchical method and the nonhierarchical method. The difference between the two methods lies in determining the number of clusters resulting from. In the hierarchical method, the determination of the number of clusters has not been determined, while in the nonhierarchical method, the number of clusters is identified at the beginning or the number of clusters generated is already known.

The hierarchical method has advantages over the nonhierarchical method. The advantages of the hierarchical method are that it is easier to study all the clusters that are formed and more informative because the hierarchical method of grouping stages is presented in the form of a dendogram or tree diagram. One of hierarchical methods, namely the ward linkage or variance method, aims to obtain clusters that have the smallest possible internal cluster variance. The method commonly used is the ward method with the average for each cluster calculated. Furthermore, in the cluster analysis there are several similarity measures measured by distance, namely Euclidean, Manhattan/city block and mahalanobis. The three distance measures have their respective advantages and disadvantages, so research is needed on the comparison of the three to get the best distance. If the lower the value of the distance between objects, the higher the similarity of objects.

The house is one of the primary human needs besides the need for clothing and food. However, housing needs are not easily fulfilled by some people. Therefore, the bank offers a solution in the form of a Home Ownership Credit (KPR) product. Home Ownership Credit (KPR) is a facility provided to buy a residence (house) using credit to a bank. In practice, there are problems that are often faced by banks in running mortgage

Journal of Theoretical and Applied Information Technology
31st January 2023. Vol.101. No 2

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

products, namely bad loans which can cause a financial crisis [2]. Therefore, the measurement of the debtor by the bank is important in knowing the debtor can make payments smoothly or not. To overcome this, the appropriate statistical method used is cluster analysis.

Each method of measuring distance has advantages and disadvantages. The advantage of Euclidean distance is that it is relatively easy to calculate but very sensitive to sample size and correlation between variables. Manhattan distance is the easiest distance measurement method to apply, but the cluster that is formed will be invalid if the data has a correlation. In Mahalanobis distance, correlations between variables that may exist are not taken into account and each variable is given the same weight, but if the data does not have a correlation then the Mahalanobis distance will be considered the same as the standardized Euclidean distance.

Badung [3] has research on cluster analysis. Research was conducted to find the best distance measure between Euclidean distance, Manhattan distance and Mahalanobis distance. The conclusion is that Mahalanobis distance is the best distance measurement method. There is a study regarding the comparison of Euclidean, Manhattan, Minkowski, Cosine, and Mahalanobis distances in cluster partitioning around medoids (PAM) analysis [4]. The results show that the Mahalanobis distance is the best distance in conducting cluster analysis. In this study, distance comparisons have not been carried out on hierarchical cluster analysis, especially the ward linkage method so that the originality of this study is to compare the distances in hierarchical cluster analysis, especially with ward linkage. Based on the originality of the research and research problems, this research was conducted to see the grouping and comparison of the results of cluster analysis using the ward linkage method with various distances (Euclidean, Manhattan and Mahalanobis) with an assessment of the 5C variables (Character, Capacity, Capital, Collateral, Economic Condition) in Bank X mortgage customers.

2. LITERATURE REVIEW

2.1. Cluster Analysis

Cluster analysis is a multivariate analysis used in grouping objects into several clusters (groups) based on the similarity of the observed variables so that objects that have the same resemblance will be in a cluster and can be compared between objects from different clusters [5]. Santoso [6] has a research that has the main purpose of cluster analysis is to group objects based on the similarity of characteristics between these objects. Objects can be in the form of products (goods and services), objects (plants and others), and people (respondents, consumers and so on). The object will be classified into one or more clusters (groups) so that objects in one cluster will have similarities with one another. The method that is often used is the agglomerative hierarchical method.

Basically there are two methods of cluster analysis, namely the hierarchical method and the non-hierarchical method. The hierarchical method is a gradual cluster analysis method. The number of clusters formed is $N-1$ by N is the number of research objects so that it will produce all possible number of clusters formed. This method will produce a tree diagram (dendogram) that describes the stages of the cluster formation process. The hierarchical method can be divided into two, namely the agglomerative process (merger) and the divisive process (division / separation).

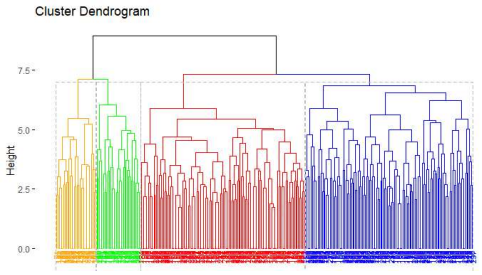

Figure 2.1 Hierarchical cluster dendogram

The agglomerative process will make each object into a cluster so that the initial number of clusters will be equal to the number of objects. Furthermore, each cluster with a high degree of similarity will be combined into a new cluster. This stage is repeated until there is only one cluster that contains all the research objects. While the divisive method is briefly the opposite of the agglomerative method. All existing objects will be combined into one cluster, then objects with a high degree of difference will be separated and become a new cluster. This step is carried out continuously until each cluster consists of only one object. This research will focus on the use of agglomerative hierarchical cluster analysis.

Journal of Theoretical and Applied Information Technology
31st January 2023. Vol.101. No 2

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

2.1.1. Grouping Method

If at the time a cluster has more than one data member, then the distance measurement method cannot be used so that a method is needed to combine clusters that have more than one member. Mattjik et al. [7] said that the grouping method in an agglomerative hierarchy is used to group objects in a structured manner based on their similarity in nature. In addition, this method is used if at the beginning of the analysis there is lack of information about the characteristics of the objects to be grouped so that the number of desired groups is not known. This research uses the ward linkage method.

2.1.1.1. Ward linkage method

The variance method aims to obtain clusters that have the smallest possible internal cluster variance. The method commonly used is the ward method with the average for each cluster calculated. At each stage, the two clusters that have the smallest increase in the ' sum of squares' in the cluster are combined [8].

Ward method is a cluster formation method based on the loss of information due to merging objects into clusters. It is measured using the total number of squared deviations in the cluster mean for each observation. Error Sum of Squares (SSE) is used as an objective function. Two objects will be combined if they have the smallest objective function among the possibilities. The objective function of the ward linkage method is presented in equation (2.1).

$$SSE = \sum_{k=1}^{v}\left(\sum_{i=1}^{n}\varepsilon_{ik}^{2} - \frac{1}{n}\left(\sum_{i=1}^{n}\varepsilon_{ik}^{2}\right)^{2}\right) \quad (2.1)$$

Information:

ε_{ij} : error value for the i-th object in the cluster k

v : number of variables

n : the number of respondents in the formed cluster

2.1.2. Distance Measurement Method

Agglomerative hierarchy method, a similarity measurement method is needed to determine the objects to be combined into a new cluster. Hair et al. [9] said that there are three methods of measuring similarity, namely correlational measures, distance measures, and association measures. The most frequently used similarity measurement method is distance measures. 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 it, and vice versa. Hair et al. [9] also said that there are several methods of measuring distance, including Euclidean distance, City-Block (Manhattan) distance, and Mahalanobis distance.

2.1.2.1. Euclidean distance

Euclidean distance is the most commonly used type of distance measurement because it is one of the easiest methods to understand and model. Euclidean distance is used to measure the distance from the data object to the center of the cluster. This method is appropriate to use to determine the closest distance between two data. Euclidean distance is the geometric distance between two data objects [10]. The Euclidean distance between two points can be calculated using equation (2.2).

$$d\left(x_i, x_j\right) = \sqrt{\sum_{z=1}^{v}\left(x_{li} - x_{lj}\right)^2} \quad (2.2)$$

Information:

$d(x_i, x_j)$ euclidean distance between object i and object j

x_{li} : value of object i in variable l

x_{lj} : jth object value in variable l

z : variable to z ,$z = 1,2,3, \ldots, v$

2.1.2.2. Manhattan distance

Manhattan distance is used to calculate the absolute difference between the coordinates of a pair of objects. Prasetyo [11] said that the Manhattan distance is very appropriate for detecting outliers in the data. Manhattan distance between two points can be calculated using equation (2.3).

$$d(x_i, x_j) = \sum_{r=1}^{v}|x_{li} - x_{lj}| \quad (2.3)$$

Information:

$d(x_i, x_j)$: manhattan distance between object i and object j

x_{li} : value of variable l on object i

x_{lj} : value of variable l on jth object

v : many data variables\

2.1.2.3. Mahalanobis distance

Mahalanobis distance is a generalization of the standardized Euclidean square distance. Seber [12] revealed that the mahalanobis distance can not only solve the problem of differences in the data scale, but also consider the effect of correlation between variables. The mahalanobis distance between two points can be calculated using equation (2.4).

Journal of Theoretical and Applied Information Technology
31st January 2023. Vol.101. No 2

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

$$D_{ij}^2 = \frac{1}{(1-r^2)}\left[\frac{(x_{i1}-x_{j1})^2}{S_1^2} - 2r\frac{(x_{i1}-x_{j1})}{S_1}\frac{(x_{i1}-x_{j1})}{S_2} + \frac{(x_{i2}-x_{j2})^2}{S_2^2}\right]$$

Information:

D_{ij}^2 : observation distance of object i and object j

S_1^2 : variance for the 1st variable

S_2^2 : variance for the 2nd variable

S_1S_2 : covariance of the 1st and 2nd variables

2.1.3. Determination of the Goodness of the Best Distance Measurement Method

In accordance with the main objective to be achieved in this study, which is to compare several distances in the ward method, calculations are carried out using a certain size to see which method gives optimal results. Standard deviation can be used to determine the homogeneity between clusters. The cluster that is formed is said to be good if it has a standard deviation value in the cluster (S_W) the minimum and the maximum standard deviation between clusters (S_B) [13]. Standard deviation formula in cluster (S_W) can be seen in equation (2.5).

$$S_W = \frac{\sum_{k=1}^{K} S_k}{K} \quad (2.5)$$

where:

S_k : standard deviation of k cluster

K : the number of clusters formed

As for the standard deviation between clusters (S_B) can use equation (2.6).

$$S_B = \sqrt{\frac{\sum_{k=1}^{K}\left(\overline{X}_k - \overline{X}\right)^2}{(K-1)}} \quad (2.6)$$

where:

$\overline{X}_k$ kth cluster mean

$\overline{X}$: the average of the entire cluster

K : the number of clusters formed

The value of the standard deviation ratio is obtained from the value of (S_W) and (S_B). The method with the largest standard deviation ratio can be considered the most optimal method [14]. The greater the value of the standard deviation ratio of a method, the better the method used [15]. The calculation of the standard deviation ratio can use the equation (2.7).

$$\text{standard deviation ratio} = \frac{S_B}{S_W} \quad (2.7)$$

where

S_W : standard deviation in cluster

S_B : standard deviation between clusters

2.2. Research variables

2.2.1. Character

Kustini [16] revealed that character is the first and very important benchmark in measuring lending to debtors. Character is a reflection of a person's self [17]. The bank will conduct an interview process between the debtor and creditor to find out whether a debtor can answer the questions asked properly and honestly or vice versa. Through the interview process, the bank will find out a debtor, including people who can pay off payments properly or not.

2.2.2. Capacity

Kustini [16] said the capacity or ability of the debtor to repay and repay the loans he borrowed is very important and is the main consideration for the bank in providing credit to debtors. In this case, the bank has an assessment in analyzing the ability to repay the credit that will be given to the debtor.

2.2.3. Capital

Loppies [18] revealed that capital is the own capital owned by prospective debtors. The greater the own capital owned, the bank will be more confident in deciding to give credit. In assessing the capital owned by the prospective debtor, the bank must analyze of the overall financial position, both for the past and for the future so that it can be seen the ability of the prospective debtor's capital in supporting the financing of the prospective debtor's business [19].

2.2.4. Collateral

Kustini [16] said collateral is very important for the bank because the presence of collateral can bind customers in the event of a

non-performing loan. The Bank will not provide financing that exceeds the value of the guarantee, except for certain financing guaranteed by certain parties.

2.2.5. Condition of Economy

Djachrab [20] revealed that the condition of economy is a condition specifically related to the economic aspect which is likely to affect the business or business where the prospective customer works in the future if there is a change [21]. This assessment is directed at the surrounding conditions that affect the prospective customer's business [22].

2.2.6. Credit Collectability

According to Bank Indonesia Regulation Number 14/15/PBI/2012 concerning the assessment of the asset quality of commercial banks, three groups of credit collectibility are determined, namely current loans, substandard loans, and bad loans.

1. Current credit is a condition when there are no arrears in principal installments, arrears in interest or overdrafts due to credit withdrawals. Current credit is also defined as the debtor making payments on the debt as well as interest on a timely basis.
2. Substandard credit is a condition when there are arrears of obligations (principal and interest installments or interest only) with in 30 working days.
3. Bad credit is a condition when there are arrears of obligations (principal and interest installments or interest only) for more than 30 working days.

Credit collectability can be used to see the performance of a bank [23]. One technique is to determine the collectibility level of the bank's debtors so that it can be used as a benchmark for the bank's performance in the future. The value of a successful company is related to good performance. The long-term goal of a company is to increase the value of the company [24].

3. RESEARCH METHODS

3.1. Data Source

The data used in this study is the primary data variable Character, Capacity, Capital, Collateral, and Condition of Economy on KPR Bank X customers. Measurement of research variables in the economic field is generally done by asking for opinions or perceptions of someone through a questionnaire [25]. Therefore, this study made a questionnaire and distributed it to the respondents. The questionnaire will be tested for validity and reliability [26]. It is critical to test instrument validity in order to obtain unbiased data (as a result of the instrument/questionnaire) [27]. The population in this study were all mortgage debtors of Bank X, one of the cities in Indonesia. The number of respondents in this study w as 100 customers. The sampling technique used was purposive sampling. Purposive sampling is a sampling technique based on certain characteristics and considerations [28]. The scale model used is the Likert scale. Fernandes et al. [29] said that the determination of the response item on the Likert scale in the form of a score, with 5 response items, namely a score of 1 = strongly disagree, 2 = disagree, 3 = neutral, 4 = agree and a score of 5 = strongly agree.

3.2. Steps

The steps taken in this study are as follows:

1. Obtain latent variable data by making a questionnaire and distributing it to respondents.
2. Scaling the data using the method of measuring the average score of the indicator. Solimun & Fernandes [27] said the average scale method is carried out by using the average scale of all indicators on each variable so that the average scale data is obtained which is the data of the relevant latent variable.
3. Grouping mortgage customers based on the assessment of the 5C variable with Cluster analysis using the ward linkage method with various distance measurements.
4. Comparing the goodness of all distance measurement methods using the standard deviation ratio.

4. RESEARCH RESULTS

4.1. Cluster Analysis Results

Cluster analysis aims to group and set diverse objects into relatively homogeneous clusters based on certain characteristics. In this study, an agglomerative hierarchical cluster analysis was carried out to classify Bank X mortgage customers based on a 5K assessment using the grouping method, namely the ward linkage method with several distance measurements, namely Euclidean distance, Manhattan distance, and Mahalanobis distance.

4.1.1. Results of Cluster Analysis Using Ward Linkage with Euclidean Distance

Journal of Theoretical and Applied Information Technology
31st January 2023. Vol.101. No 2

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

The Euclidean distance method will measure the distance between objects using equation (2.2), then grouped using the Ward method, by taking the smallest EES value. The dendogram of the Euclidean distance using the Ward method of two clusters is presented in Figure 4.1.

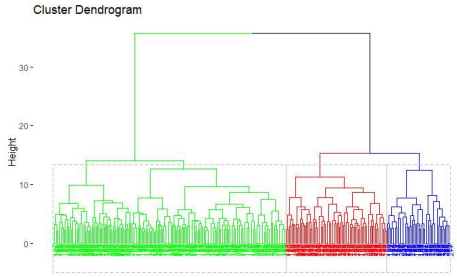

Figure 4.1. Euclidean Distance Dendogram With Ward Method

Based on Figure 4.1, it can be seen that using the Euclidean distance with the Ward method on customer data for KPR Bank X based on the 5C assessment can be divided into 3 clusters. Cluster one is green, cluster two is red, and cluster three is blue. Members in each cluster can be seen in Table 4.1.

Table 4.1. Many Members Of Three Clusters Use Euclidean Distance With The Ward Method

Cluster to-	Customer Code	Number of Cluster Members
Cluster 1	5, 8, 9, 10, 12, 15, 16, 18, 22, 26, 27, 29, 32, 36, 37, 39, 40, 46, 48, 52, 58, 67, 78, 80, 81, 84, 87, 90, 91, 94, 99 100	32
Cluster 2	1, 17, 20, 21, 25, 34, 35, 43, 44, 45, 49, 54, 57, 59, 60, 66, 68, 71, 73, 76, 82, 85	22
Cluster 3	2, 3, 4, 6, 7, 11, 13, 14, 19, 23, 24, 28, 30, 31, 33, 38, 41, 42, 47, 50, 51, 53, 55, 56, 61, 62, 63, 64, 65, 69, 70, 72, 74, 75, 77, 79, 83, 86, 88, 89, 92, 93, 95, 96, 97, 98	46

4.1.2. Results of Cluster Analysis Using Ward Linkage with Manhattan Distance

The Manhattan distance method will measure the distance between objects using equation (2.3), then grouped using the Ward method , using the smallest SSE value. The dendogram of the Euclidean distance using the Ward method of two clusters is presented in Figure 4.2.

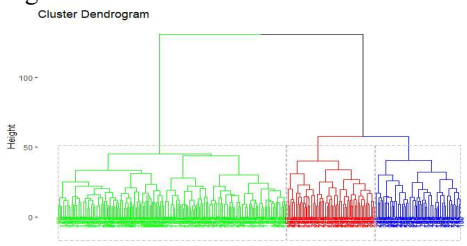

Figure 4.2. Manhattan Distance Dendogram Using Ward Method

Based on Figure 4. 2 , it can be seen that using the Manhattan distance with the Ward method on the data of KPR Bank X creditors based on the 5C assessment, 3 clusters are formed. The number of members from each cluster can be seen in Table 4.2.

Table 4.2. Multiple Members Of Three Clusters Using Manhattan Distance With Ward Method

Cluster to-	Customer Code	Number of Cluster Members
Cluster 1	1, 5, 8, 12, 17, 20, 21, 25, 34, 35, 36, 43, 44, 45 , 52, 54, 57, 58, 59, 60, 66, 68, 71, 73, 76, 82, 85, 90 , 94, 99	29
Cluster 2	8, 9, 10, 12, 15, 16, 18, 22, 26, 27, 29, 32, 36, 37, 39, 40, 46, 48, 52, 58, 67, 78, 80, 81, 84, 87, 91	27
Cluster 3	2, 3, 4, 6, 7, 11, 13, 14, 19, 24, 28, 30, 31, 33 , 38, 41, 42, 47, 50, 51, 53, 55, 56, 61, 62, 63, 64, 65, 69, 70, 72, 74, 75, 77, 79, 83, 86, 88, 92, 93, 95, 96, 97, 98	44

Journal of Theoretical and Applied Information Technology
31st January 2023. Vol.101. No 2

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

4.1.3. Cluster Analysis Results Using Ward Linkage with Mahalanobis Distance

Mahalanobis Method distance will measure the distance between objects using equation (2.4), then grouped using the Ward method, using the smallest ESS value. Mahalanobis distance dendogram result with Ward's method of two clusters is presented in Figure 4.3.

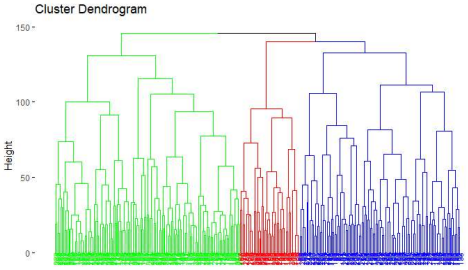

Figure 4.3. Mahalanobis Distance Dendogram With Ward method

Based on Figure 4.3, it can be seen that using the Mahalanobis distance with the Ward method on the data of Bank X mortgage creditors based on the 5C assessment, three clusters are formed. The number of members in each cluster can be seen in table 4.3.

Table 4.3 Many Members In 3 Clusters With Mahalanobis Distance And Ward Method

Cluster to-	Customer Code	Number of Cluster Members
Cluster 1	9, 10, 15, 16, 18, 22, 23, 26, 27, 29, 32, 37, 39, 40, 46, 48, 49, 67, 78, 80, 81, 84, 87, 89, 91, 100	26
Cluster 2	1, 5, 8, 12, 17, 20, 21, 25, 34, 35, 36, 43, 44, 45 , 52, 54, 57, 58, 59, 60, 66, 68, 71, 73, 76, 82, 85, 90 , 94, 99	30
Cluster 3	2, 3, 4, 6, 7, 11, 13, 14, 19, 24, 28, 30, 31, 33 , 38, 41, 42, 47, 50, 51, 53, 55, 56, 61, 62, 63, 64, 65, 69, 70, 72, 74, 75, 77, 79, 83, 86, 88, 92, 93, 95, 96, 97, 98	44

4.2. Best Clustering Results

Clustering results is useful to find out how well the performance of all methods used to group research objects is. A cluster or group is said to be good if it has high homogeneity within the cluster and high heterogeneity between clusters. The way to know the homogeneity within the cluster is to calculate the standard deviation within the group (Sw) while to determine the heterogeneity between clusters using the standard deviation between groups (Sb). So that the best grouping results have the maximum standard deviation between groups (Sb) and the maximum standard deviation within the group (Sw). The results of the standard deviation ratio are presented in Table 4.4.

Table 4.4. Standard Deviation Ratio Within Clusters And Between Clusters

Distance Measurement Method	Standard Deviation		Standard Deviation Ratio
	S_w	S_b	
Euclidean Distance	0.990	0.365	0.368
Manhattan Distance	0.967	0.377	0.390
Mahalanobis Distance	0.958	0.624	**0.652**

Based on Table 4.4, it is known that the method that produces the largest standard deviation ratio between groups (Sb) and standard deviation within groups (Sw) is the Mahalanobis distance method with the Ward method of 0.652. The greater the value of the standard deviation ratio of a cluster method , the better the performance of the cluster method in grouping objects. Thus, the performance of the Mahalanobis distance method can be said to be better than the performance of other distances in the grouping of mortgage customers at Bank X in 2021 based on the 5C assessment.

4.3. Interpretation of Cluster Results Using Mahalanobis Distance with Ward method

In subchapter 4.2, Mahalanobis distance with Ward method is the most optimal hierarchical cluster method with 3 clusters. Each cluster has different characteristics. Calculation of the average value of each variable in each cluster as on Table 4.5.

Journal of Theoretical and Applied Information Technology
31st January 2023. Vol.101. No 2

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Table 4.5. Average Variables In Each Cluster

Variable	Cluster 1	Cluster 2	Cluster 3
Character (Y_1)	2.8513	2.8324	2.8365
Capacity (Y_2)	2.6112	2.3841	2.9946
Capital (Y_3)	2.8647	2.1889	3.0033
Wealth (Y_4)	2.6745	2.6710	2.7580
Condition (Y_5)	2.8556	3.0097	2.8671
Average	2.805	2,681	2,905

Characteristics in each cluster can be shown by calculating the average of each variable in each cluster. Based on Table 4.23, it can be seen that the overall average value of all 5C assessment variables is in sufficient condition (mean value > 2.5). The data in this study uses a Likert scale of 1 to 5, meaning that with an average value of > 2.5, Bank X mortgage customers can be considered good in meeting the 5C assessment.

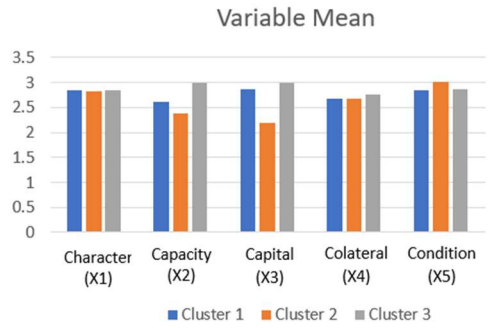

Figure 4.4. Average Of 5C Assessment Variables From Clustering Results Using Mahalanobis Distance With Ward 'S Method

Based on Figure 4.4, it can be seen that cluster 1 has a slightly higher character rating, but is slightly lower in capacity and wealth compared to other clusters. So that customers in cluster 1 can be considered to have the best character or personality when compared to customers in other clusters , but the value of the customer's capacity or ability to pay installments and assets owned by customers is lower than customers in cluster 3 but higher if compared to customers in cluster 2.

Customers who are in cluster two have a condition assessment level and the highest compliance among the existing clusters. But the level of character, capacity, capital , and wealth is at the lowest level. This means that customers in cluster two have good economic conditions and compliance with paying installments. The assessment of the personality and assets owned by customers in cluster 2 is slightly lower, while the ability of customers to pay installments and external conditions owned by customers in cluster 2 is the lowest with a considerable difference compared to customers in other clusters. Cluster 2 can be considered as a cluster with members with the lowest 5C rating because cluster 2 has the lowest value on four of the five variables used.

In cluster three, the capacity value, capital and wealth is at the highest level, while in the other assessments it is slightly lower than the other clusters. The conclusion in cluster three is that customers who are members of cluster three have the highest level of ability to pay installments, financial conditions, and external conditions when compared to members in other clusters. Cluster three has a slightly lower assessment of personality, external conditions, and compliance in paying installments but not the lowest among the three clusters formed. So that cluster three can be considered as a cluster with members who have the best 5C rating compared to members in other clusters.

The research was conducted using the ward method with several distances. This research is different from previous research which has been explained in subchapter 1 in that there are several studies that perform cluster analysis by comparing several distances but not using the linkage method, especially using the ward method. Previous research said that the mahalanobis distance was the best distance in conducting cluster analysis. This is supported by the results of this study, namely the best distance results are using the mahalanobis distance although the linkage method used is different.

In the results of clustering using the ward linkage method with a distance of mahalanobis, there are customers belonging to the low cluster so that the bank needs to take countermeasures against customers belonging to the low cluster. One of the countermeasures is to extend the time period so that debtors are able to pay their debts because customers in the low cluster do not yet have sufficient confidence to be able to complete loans on time so that with these efforts it is hoped that customers in this cluster are able to pay credit. Customers belonging to the current cluster have fairly good economic conditions and have sufficient confidence to be able to complete loans on time. Therefore, an effort that can be done by banks is that banks need to review the contents of the credit agreement (reconditioning). Reconditioning is an attempt to make changes to part or all of the terms of the agreement, not

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

limited to changes in the installment schedule and or credit period or without converting all or part of the credit. A review of the contents of the credit agreement is carried out so that things that cause customers are not confident enough to pay on time can be resolved [30]. Another effort that can be done by banks is restructuring. Based on the Circular Letter of Bank Indonesia (SEBI) No. 23/12/BPPP dated February 28, 1991, restructuring was carried out by changing the terms of the credit agreement in the form of granting additional credit, or converting the entire portion of the loan.

5. CONCLUSIONS AND SUGGESTIONS

5.1. Conclusion

Based on the results of the analysis and discussion that has been done, it can be concluded that:

1. Mahalanobis distance with the ward method produces 3 clusters with cluster 1 as many as 26 customers, cluster 2 as many as 30 customers, and cluster 3 as many as 44 customers. Each cluster have different characteristics. Overall, cluster 1 is a medium cluster, cluster 2 is a cluster low. Meanwhile , cluster 3 has a high cluster.
2. The best performance of the clustering method in measuring distance and grouping mortgage customers at Bank X is Mahalanobis distance with the Ward method.
3. In accordance with the problem with bank mortgages, customers classified as low credit can be overcome by extending the time period so that the debtor is able to pay his debts because customers in the low cluster do not have sufficient confidence to be able to complete loans on time and banks need to review the contents of the credit agreement (reconditioning).

5.2. Suggestion

Some suggestions that can be given based on the results of this study are as follows:

5.2.1. Theoretical Suggestions

1. If the research is to be conducted using cluster analysis, it can be used in combination with the Mahalanobis distance measurement method with the Ward method because the results of the study show that this combination is the best among others.
2. The limitation of this research is to use the ward method, further research can be done using other linkage methods.
3. For further research, comparisons can be made between the distances from one another by using other validity measures or indices such as the Shilloutte index and Davies Bouldin.

5.2.2. Practical Advice

1. Bank X needs to undertake several countermeasures in dealing with customers belonging to the medium to low cluster, namely rescheduling credit payments (rescheduling), reconditioning and restructuring.

REFERENCES

[1] Narimawati, U., "Metodologi Penelitian Kualitatif dan Kuantitatif, teori dan Aplikasi", Agung Media, 2008.
[2] Hakim, W., & Fernandes, A., "Moderation effect of organizational citizenship behavior on the performance of lecturers," Journal of Organizational Change Management, 2017.
[3] Badung, N. M. A. A. 2021. "Perbandingan Ukuran Jarak pada Integrated Cluster Analysis dengan Multiple Discriminant Analysis pada Konsumen KPR Bank X", Thesis, 2021
[4] Mohammed, N. N., & Abdulazeez, A. M, "Evaluation of partitioning around medoids algorithm with various distances on microarray data", In 2017 IEEE International Conference on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData), pp. 1011-1016, 2017
[5] Supranto, J, "Analisis Multivariat Arti dan Interpretasi", Rineka Cipta, 2004.
[6] Santoso, S, "Seri solusi bisnis berbasis TI: Menggunakan SPSS untuk statistik multivariat", Elex Media Komputindo, 2006.
[7] Mattjik, A., dkk, "Aplikasi Analisis Peubah Ganda",Depdiknas, 2002.
[8] Simamora, B., "Analisis multivariat pemasaran", Gramedia Pustaka Utama, 2005.
[9] Hair, Joseph F., Black, William C., Babin, Barry J. dan Anderson, Rolph E,

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

"Multivariate Data Analysis (Seventh edition)", Boston: Pearson, 2010.
[10] Johnson, R. A. dan Wichern D. W., "Applied Multivariate Statistical Analysis (Sixth edition)"., Pearson Education, Inc, 2007.
[11] Prasetyo, E., "Data mining konsep dan aplikasi menggunakan matlab", Andi, 2012
[12] Seber, G. A. F., "Multivariate Obsevations" John Wiley & Sons, Inc, 2004.
[13] Hardle, Wolfgang K., dan Simar, Leopold, "Applied Multivariate Statistical Analsis", Springe-Verlag, 2003
[14] Hidayat, M. F., & Achmad, R. F. A., "Estimation of Truncated Spline Function in Non-parametric Path Analysis Based on Weighted Least Square (WLS)", In IOP Conference Series: Materials Science and Engineering, vol. 546, no. 5, pp. 052027, 2019. DOI: 10.1088/1757-899X/546/5/052027
[15] Fernandes, A. A. R., "Investigation the mediating variable: What is necessary?(case study in management research", International Journal of Law and Management, vol. 59, no. 6, pp. 1059-1067, 2017. DOI: 10.1108/IJLMA-09-2016-0077
[16] Kustini, F. D., "Analisis Faktor 3R dan 5C dalam Pemberian KPR", Jurnal Ekonologi Ilmu Manajemen, vol. 5, no. 2, pp.231-241, 2017.
[17] Musadieq, M., Nurjannah, N., Raharjo, K., Solimun, S., & Fernandes, A. A. R., "The mediating effect of work motivation on the influence of job design and organizational culture against HR performance", Journal of Management Development, vol. 37, no. 6, pp. 452-469, 2018. DOI: 10.1108/JMD-07-2017-0239
[18] Loppies, L. S., Essomar, M. J., & Rikumahu, B. F., "Analisis Pengaruh 5c (Character, Capacity, Capital, Collateral, Condition Of Economic) Terhadap Keputusan Kredit Di Pt. Bpr Modern Express Ambon", Soso-Q: Jurnal Manajemen, vol. 9, no.1, pp.88-107, 2021.
[19] Fernandes, A. A. R., "Moderating effects orientation and innovation strategy on the effect of uncertainty on the performance of business environment". International Journal of Law and Management, vol. 59, no. 6, pp. 1211-1219. 2017. DOI: 10.1108/IJLMA-10-2016-0088
[20] Djachrab, I. R., Faktor 5C dalam Pembiayaan Kepemilikan Rumah di Bank BJB Syariah Kantor Cabang Pembantu Rancaekek", Jurnal Maps (Manajemen Perbankan Syariah), vol. 4, no. 2, pp. 36-46, 2021.
[21] Sumardi, S., & Fernandes, A. A. R., "The influence of quality management on organization performance: service quality and product characteristics as a medium", Property Management, vol. 38, no. 3, pp. 383-403, 2020. DOI: 10.1108/PM-10-2019-0060
[22] Sri, M., & Solimun, S., "The relationship between audit quality and risk taking toward value creation in Indonesia", Journal of Accounting in Emerging Economies, vol. 9 no. 2, pp. 251-267, 2019, DOI: 10.1108/JAEE-11-2017-0110
[23] Raharjo, K., Nurjannah, N., Solimun, S., & Fernandes, A. A. R., "The influence of organizational culture and job design on job commitment and human resource performance", Journal of Organizational Change Management, vol. 31, no. 7, pp. 1346-1367, 2018. DOI: 10.1108/JOCM-07-2017-0286
[24] Purbawangsa, I. B. A., Solimun, S., Fernandes, A. A. R., & Rahayu, S. M., "Corporate governance, corporate profitability toward corporate social responsibility disclosure and corporate value (comparative study in Indonesia, China and India stock exchange in 2013-2016)", Social Responsibility Journal, vol. 16, no. 7, pp. 983-999, 2019. DOI: 10.1108/SRJ-08-2017-0160
[25] Fernandes, S., & Rinaldo, A. A. R. A. A., "The mediating effect of service quality and organizational commitment on the effect of management process alignment on higher education performance in Makassar, Indonesia", Journal of Organizational Change Management, vol. 31, no. 2, pp. 410-425, 2018. DOI: 10.1108/JOCM-11-2016-0247.
[26] Kirono, I., Armanu, A., Hadiwidjojo, D., & Solimun, S., "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, vol. 36, no. 8, pp. 1301-1317., 2017. DOI: 10.1108/IJQRM-11-2017-02462019.
[27] Solimun; Fernandes, Adji Achmad Rinaldo., "Investigation of Instrument Validity: Investigate the Consistency between

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Criterion and Unidimensional in Instrument Validity (Case Study in Management Research)". Int'l JL & Mgmt., vol. 59, pp. 1203, 2017.

[28] Fernandes, A. A. R., Budiantara, I. N. I., Otok, B. W., & Suhartono, "Reproducing Kernel Hilbert space for penalized regression multi-predictors: Case in longitudinal data", International Journal of Mathematical Analysis", vol. 8, no. 40, pp. 1951-1961, 2014. DOI: 10.12988/ijma.2014.47212

[29] Fernandes, A. A. R., Hutahayan, B., Arisoesilaningsih, E., Yanti, I., Astuti, A. B., & Amaliana, L., "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, vol. 546, no. 5, pp 052024, 2019. DOI: 10.1088/1757-899X/546/5/052024.

[30] Fernandes, A. A. R., & Taba, I. M., "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, vol. 10, no. 1, pp. 58-72, 2018. DOI: 10.1108/JSTPM-05-2017-0019