

# Multiple Discriminant Analysis Altman Z-Score, Multiple Discriminant Analysis Stepwise and K-Means Cluster for Classification of Financial Distress Status in Manufacturing Companies Listed on the Indonesia Stock Exchange in 2019

Hazrina Ishmah\*, Solimun, Maria Bernadetha Theresia Mitakda

Department of Statistics, Faculty of Mathematics and Natural Science,  
Brawijaya University, Malang, 65145, Indonesia

\*Corresponding Email: [hazrina\\_30@student.ub.ac.id](mailto:hazrina_30@student.ub.ac.id)

## ABSTRACT

This study uses the MDA (Multiple Discriminant Analysis) Altman Z-Score to predict the status of financial distress in manufacturing companies listed on the Indonesia Stock Exchange in 2019. MDA Stepwise model is used to prove that the variables used in the MDA Altman Z-Score method are the best variables for predicting financial distress status. MDA Altman Z-Score uses five variables from financial ratios. Variables used in Altman Z-Score are working capital/total assets, retained earnings/total assets, earnings before interest and taxes/total assets, market value equity/book value of total liabilities and sales/total assets. The variables used in MDA Stepwise are 38 financial ratios and validate that the MDA Altman Z-Score is appropriate in classifying manufacturing companies experiencing financial distress in 2019 using the K-Means cluster. The results obtained for the best prediction of financial distress status using MDA Stepwise seen from the highest accuracy value (84.54%) and significant variables in predicting financial distress status are capital market to book value of debt, sales/work capital, and sales/current assets variables. The best classification for manufacturing companies if they are classified into 3 groups, namely the group not experiencing financial distress, gray area and experiencing financial distress. The category of the grouping of companies resulted in 73 companies experiencing financial distress, one company was in the gray area and nine companies did not experience financial distress.

**Keywords:** Financial distress, K-means cluster, MDA Altman Z-score, MDA stepwise.

## 1. INTRODUCTION

Multiple Discriminant Analysis (MDA) aims to group objects into two or more classes. The grouping is mutually exclusive in the sense that if object A has entered group 1, then it is impossible to become group 2 [1]. MDA has been widely applied to social and economic research and one that is often applied to economic research is the Altman Z-Score MDA.

One of the studies that applied the Altman Z-score method was research used the Z-score model to identify two years of financial difficulties before bankruptcy[2]. This study is also supported by another study, which are

predicted a company's bankruptcy status with the Z-score model has an accuracy of 95% for one year and 85% for two years before the bankruptcy[3]. It is not too surprising because since begin of 1984 Altman has research various countries and managed to identify 1-2 years before the financial distress. Research in three different periods, the Altman Z-score model had different accuracy for predicting bankruptcy[4]. This indicates that the Altman Z-score model is not necessarily suitable for use in the current period compared to when it was applied in previous periods so that this model turns out to be unreliable in predicting corporate bankruptcy.

The variable used to predict the status of financial difficulties in the company is in the form of financial ratios. Many studies for predicting financial distress status use other financial ratios, one of which uses 33 financial ratios to determine the prediction of the company's difficulty status[5]. The MDA Altman Z-Score has similarities with the K-Means cluster for divided objects into groups based on the similarity of object properties in one group and differences in object properties between groups. Cluster analysis is a multivariate technique that aims to classify an object into different groups. Objects that have divided on one cluster have relatively the same distance proximity[6].

This study uses the MDA Altman Z-Score to predict the status of financial distress in manufacturing companies listed on the Indonesia Stock Exchange in 2019 into three groups, namely those experiencing financial distress, gray areas, and those not experiencing financial distress. Research Altman Z-score model is more suitable for classifying companies experiencing financial difficulties in manufacturing companies[7]. This study uses MDA Stepwise to get the best variables used in predicting the status of financial difficulties. The variables used in MDA Stepwise are 38 financial ratios. Stepwise model is used to prove that the variables used in the MDA Altman Z-Score method are the best variables for predicting financial distress status. The purpose of discriminant analysis is to classify objects and find out whether an object belongs to one group or another, K-Means cluster for an alternative in classifying manufacturing companies in Indonesia in 2019.

## 2. METHOD

The variables used are financial ratios. MDA Altman, the variables used are: financial distress ( $Y$ ), working capital/total assets ( $X_1$ ), retained earnings/total assets ( $X_2$ ), earning before interest and taxes/total assets ( $X_3$ ), market value equity/book value of total liabilities ( $X_4$ ) and sales/total assets ( $X_5$ ). MDA Stepwise uses more variables as many as 38 financial ratios used are: financial distress ( $Y$ ), working capital/total assets ( $X_1$ ), retained earnings/total assets ( $X_2$ ), earning before interest and taxes/total assets ( $X_3$ ), market value equity/book value of total liabilities ( $X_4$ ), sales/total assets ( $X_5$ ), net income/sales ( $X_6$ ), net income/total assets ( $X_7$ ), net income/equity ( $X_8$ ), total liabilities/total assets ( $X_9$ ), current liabilities/total assets ( $X_{10}$ ), current liabilities/total liabilities ( $X_{11}$ ), notes payable/total assets ( $X_{12}$ ), notes payable/total liabilities ( $X_{13}$ ), long-term debt/total assets ( $X_{14}$ ), equity/total assets ( $X_{15}$ ), long-term debt/equity ( $X_{16}$ ), current assets/current liabilities ( $X_{17}$ ), (current assets-inventory)/current liabilities ( $X_{18}$ ), current assets/total assets ( $X_{19}$ ), net fixed assets/total assets ( $X_{20}$ ), cash/current assets ( $X_{21}$ ), cash/total assets ( $X_{22}$ ), sales-growth ( $X_{23}$ ), net income/total assets-growth ( $X_{24}$ ), cost of goods sold/ inventory ( $X_{25}$ ), sales/account receivable

( $X_{26}$ ), account receivable/total assets ( $X_{27}$ ), sales/work capital ( $X_{28}$ ), sales/current assets ( $X_{29}$ ), account receivable/inventory ( $X_{30}$ ), (account receivable+ inventory)/ total assets ( $X_{31}$ ), cost of goods sold/sales ( $X_{32}$ ), sales general administration expense/sales ( $X_{33}$ ), (cost of goods sold+sales general administration expense)/sales ( $X_{34}$ ), earnings before interest and taxes/interest expense ( $X_{35}$ ), long-term debt/sales ( $X_{36}$ ), interest expense/sales ( $X_{37}$ ) and account payable/sales ( $X_{38}$ ).

### 2.1. Discriminant Analysis

The discriminant analysis aims to classify an individual into mutually exclusive/disjoint and exhaustive groups based on a number of explanatory variables. The purpose of this analysis is to find out the differences between groups [8].

Discriminant analysis model:

$$Z = b_0 + b_1X_{i1} + b_2X_{i2} + \dots + b_pX_p$$

$$Z_i = b_0 + \sum_{j=1}^p b_jX_{ij},$$

where:

- $Z_i$  : discriminant score of discriminant function I,
- $X_{ij}$  : the value of the i-th observation on the j-predictor variable,
- $i$  : number of object,
- $j$  : 1, ..., p ,
- $p$  : number of predictor,
- $b_0$  : intercept,
- $b_j$  : discriminant coefficient for each j-predictor predictor.

### 2.2. Multiple Discriminant Analysis (MDA) Altman Z-Score

Multiple Discriminant Analysis (MDA) Altman Z-Score was developed by Edward Altman in 1968. Edward I. Altman developed a multivariate Z-score model for identifying several kinds of financial ratios that considered the most important value in determining an event. Multiple Discriminant Analysis compared to financial ratio analysis are multivariate so can describe performance of all financial ratios well in a simple and predictive linear model, where univariate ratio analysis is difficult to provide a good explanation of which financial ratios explain well in assessing financial performances[10].

Altman Z-score model for manufacturing companies[11]:

$$Z = .012X_1 + .014X_2 + .033X_3 + .006X_4 + .999X_5$$

where  $Z$  : discriminant score total,  $X_1$  : working capital/total assets,  $X_2$  : retained earnings/total assets,  $X_3$  : earning before interest and taxes/total assets,  $X_4$  : market value equity/book value of total liabilities,  $X_5$  : sales/total assets.

There are point points for determining the z value that can explain whether the company will experience financial difficulties or not in the future [11].

The points of determination are divided into three categories:

- a) If the value  $Z < 1.81$  so companies are experiencing financial distress
- b) If the value  $1.81 < Z < 2.99$  so companies are grey area
- c) If the value  $Z > 2.99$  so companies are not experiencing financial distress

### 2.3. Multiple Discriminant Analysis (MDA) Stepwise

The stepwise method is a combination of the forward and backward methods. In the stepwise model, at first none of the variables are included in the discriminant function, then one variable is added to the discriminant function. Variables can be added or removed from the discriminant function until we get the best function.

The assumptions that must be accepted in the discriminant model are:

1. Data comes from the multivariate normal distribution. Testing the assumption of the multivariate normal distribution is important to test the significance of the variables and discriminant functions. The assumption test can use the Mahalanobis distance. Violation of assumptions usually results in a low level of classification accuracy.
2. The variance matrix of all predictor variables must be the same. The test is carried out to test the homogeneity of the variance matrix using Box's M test. The assumption that must be fulfilled is that the group covariance matrices of each variable are homogeneous.

Hypothesis:

$H_0$ : The diversification matrix group is homogeneous

$H_1$ : The diversification matrix group is not homogeneous

If significant score  $< \alpha = 0.05$ ,  $H_0$  rejected.

3. The data does not contain multicollinearity. Multicollinearity states that the predictor variables in the model must be independent of each other. To find out whether there is a correlation between variables

or not, one way to detect it is by looking at the value of the Variance Inflation Factor (VIF). VIF is a measure of the total diversity of one variable that can be explained by the diversity of other predictor variables. If the VIF value is more than 10, it is sufficient evidence to state that there is multicollinearity in the data to be studied.

4. The data does not contain outliers.

Outlier data detection is required before performing the analysis with MDA Stepwise. An outlier is an object that is very different from other objects. It can be described as an observation that is significantly biased (the expected value of the estimator is not the same as the population parameter), does not represent the general population, and there is undersampling (lack of sample). Outlier detection in multivariate using the mahalanobis distance.

Wilk's Lambda is an inverse quality criterion. In order to analyze the dissimilarity of the groups multivariate. Wilk's Lambda is calculated:

$$\Lambda = \prod_{k=1}^K \frac{1}{1 + \gamma k}$$

where:

$\Lambda$  = Wilk's lambda score,

$K$  = number of discriminant function,

$k$  = number of group,

$\gamma k$  = eigenvalue of the k-th function.

Hypothesis:

$$H_0: \mu_1 = \mu_2 = \mu_3$$

(The groups are not different from each other)

$$H_1: \mu_1 \neq \mu_2 \neq \mu_3 \text{ (There are different groups)}$$

If significant score  $< \alpha = 0.05$ ,  $H_0$  rejected.

Determine whether the discriminant analysis used is valid or not by calculating the Cpro and Cmax values. Cpro is the probability of achieving accuracy in the grouping of samples. Cmax is a measurement of validity based on the largest number of samples.

Cpro calculated:

$$C_{pro} = [p^2 + (1 - p)^2] \times 100\%$$

Cmax calculated:

$$C_{max} = (n_{max} / N) \times 100\%$$

### 2.4. K-Means Cluster

K-means cluster is one of the non-hierarchical clustering methods usually used because the process is simpler and faster. The k-means algorithm will divide objects into groups, where each group has a middle value (centroid). K-means randomly determines the middle value of the group, then using the Euclidean distance approach, the distance between the object and the middle value is calculated, then the object will be placed into the closest group calculated from the midpoint of the group. The process of determining the middle value and placing objects is carried out until the middle value converges (the mean of all groups does not change anymore).

Silhouette coefficient validation is a validation measure based on internal criteria. This method will evaluate the placement of each object in each cluster by comparing the average distance of objects in one cluster and the distance between objects in different clusters [12].

## 3. RESULT AND DISCUSSION

### 3.1. MDA Altman Z-Score

Table 1. Z-Score MDA Altman Z-Score

No	Z-Score	Response
1	1.555014e+01	2
2	3.090037e-01	0
3	2.610227e+00	1
4	70100164e+01	2
5	7.599532e-01	0
⋮	⋮	⋮
90	7.915588e+10	2

Based on the analysis using the MDA Altman Z-Score method as shown in Table 1, it is found that from 90 manufacturing companies, 66 companies are experiencing financial distress, 5 companies are in the gray area and 19 companies are not experiencing financial distress. The results of measuring the level of company classification using the MDA Altman Z-Score method using the press q statistic obtained a value of 45 with  $\alpha = 0.05$  and  $df = 2$ , then the chi-square value = 5.991. Because the statistical value is  $45 > 5.991$ , the classification of companies in 3 categories using the discriminant function is accurate. The level of accuracy with the MDA Altman Z-Score method obtained a value of 81.11%.

### 3.2. MDA Stepwise

#### 3.2.1. Testing Multivariate Normality

The result for the value of the mahalanobis distance can be seen in Table 2.

Table 2. Mahalanobis Distance

n	Mahalanobis Distance
1	6.1245
2	8.3023
3	61.853
4	11.761
5	40.142
⋮	⋮
90	10.176

Value of the Mahalanobis distance  $50\% <$  the chi-square value (94.44%) then accept  $H_0$  which means data on manufacturing companies listed on the Indonesia Stock Exchange in 2019 has a multivariate normal distribution.

#### 3.2.2. Testing Homogenous Within-Group

From the output using the R software, Box's M test results obtained  $p\text{-value} = 2.2e-16 < \alpha = 0.05$ . concluded that the diversification matrix is not homogeneous. This violates the discriminant function but can be done even though the assumption of homogeneity of the variance matrix is not met with the condition that the data does not have outliers [8].

#### 3.2.3. Detection Outlier

Outlier detection is done by using the Mahalanobis distance using the R software, the limit value is 84.3247. From checking the outlier value, there are 7 data containing outliers, then the 7 data are removed.

#### 3.2.4. Testing Non Multi-collinierity

Table 3. VIF Score

Var	VIF	Var	VIF	Var	VIF
X1	144.3	X14	49.54	X27	9.006
X2	24.65	X15	165.9	X28	1.292
X3	5.569	X16	2.099	X29	14.8
X4	93.85	X17	117.9	X30	26.12
X5	19.45	X18	160.9	X31	21.85
X6	149.4	X19	118.4	X32	20.94
X7	1569	X20	14.30	X33	8011
X8	2.168	X21	19.42	X35	1.04
X9	82.89	X22	7.27	X36	7990
X10	89.35	X23	1.495	X37	1.55
X11	158	X24	1.596	X38	5.322
X12	143.8	X25	34.69		
X13	137.9	X26	1.525		

From Table 3, it can be seen that there are values that are more than 10, so there is sufficient evidence that there is multicollinearity in the data. To eliminate multicollinearity, variable deletion is done. After deleting the variables, there are 26 variables that do contain multicollinearity and deleting 26 variables because it has a VIF value > 10.

3.2.5. Wilk's Lambda Test

Testing using R software with a significance level of 5% and 1% produces significant variables or has an average difference between groups, namely the X4 variable of 0.000, X28 of 0.000997, and X29 of 0.004705. The four variables have a value < significance level, which means reject H0. So there is a significant difference between the groups. The variable is the same as the variable selected in the discriminant function.

3.2.6. MDA Stepwise Model

MDA Stepwise only produce 2 predictor variables used for MDA, there are variable X4, X28 and X29. So the model formed is as follows:

$$D1 = 0.000313X_4 + 0.00124X_{28} + 0.4018X_{29}$$

$$D2 = 0.0000308X_4 - 0.0196X_{28} - 0.1373X_{29}$$

From discriminant function, the best variables for predicting financial distress status are X4 (the ratio of capital market value to book value of debt), X28 (sales/work capital), and X29 (sales / current assets).

3.2.7. Validity Discriminant Analysis

Calculation using R software obtained a Cpro value of 61.03934 which means the chance of achieving accuracy in grouping samples is 61.03% and Cmax of 73.49398 which means the validity value based on the number of samples is 73.49% with a hit ratio value (grouping accuracy) of 85.54%. The classification of the sample data is accurate because the hit ratio value is greater than Cpro and Cmax (Cpro < hit ratio > Cmax) which is 61.03% < 85.54% > 73.49%.

3.3. K-Means Cluster

3.3.1. KMO Score

From Table 4, the average KMO value for all variables is 0.51. It can be concluded that the sample is appropriate in presenting the population and the data is suitable for the k-means cluster.

Table 4. KMO Score

Var	KMO Score	Var	KMO Score	Var	KMO Score
X2	0.77	X16	0.28	X28	0.19
X3	0.68	X17	0.62	X29	0.28
X4	0.54	X20	0.35	X30	0.43
X5	0.38	X21	0.47	X31	0.54
X6	0.64	X22	0.45	X32	0.60
X8	0.45	X23	0.44	X35	0.40
X9	0.49	X24	0.62	X37	0.59
X10	0.43	X26	0.47	X38	0.87
X13	0.25	X27	0.65		

3.3.2. K-Means Cluster Plot

The results of the R software output using the k-means cluster found that there were 3 groups.

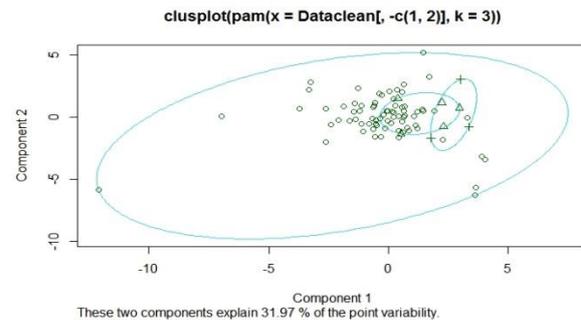


Figure 1 K-Means Cluster Plot.

Based on the results of clustering and from Figure 1, it was found that the group for companies experiencing financial distress was 68, companies in the gray area were 8, and companies not experiencing financial distress was 7. The score of the K-Means accuracy level for the classification of manufacturing companies with R software is 81.93%.

3.2.3. Validity Cluster Silhouette Coefficient

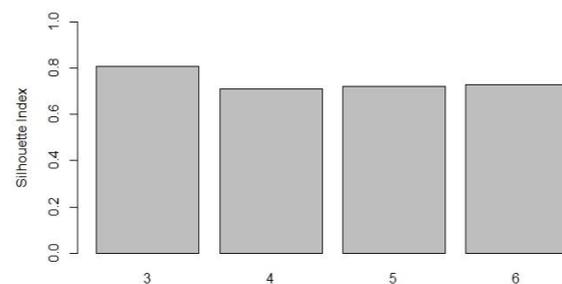


Figure 2 Silhouette Coefficient Index.

Figure 2 showed that the highest result when grouped into 3 groups, resulted in a value of 0.8090.

The accuracy level of classification accuracy resulted in a value of 81.93%. The results of the accuracy of the classification accuracy using the MDA Altman Z-Score, MDA Stepwise, and K-means Cluster methods were found that the highest value in the MDA Stepwise method was 84.54%.

#### 4. CONCLUSION

This study discusses the prediction of the company's financial distress status and the classification accuracy of manufacturing companies listed on the Indonesian stock exchange in 2019 using the MDA Altman Z-Score, MDA Stepwise, and K-Means Cluster methods. The results obtained for the best prediction of financial distress status using MDA Stepwise seen from the highest accuracy value and significant variables in predicting financial distress status are capital market to book value of debt, sales/work capital, and sales/current assets variables. The best classification for manufacturing companies if they are classified into 3 groups, namely the group not experiencing financial distress, gray area and experiencing financial distress.

Things that can be used as suggestions for further researchers based on the results of the analysis is to be able to use other discriminant methods other than MDA Stepwise. It is necessary to check more on the sample of companies used, preferably in the same company category or operating in the same field.

#### ACKNOWLEDGMENTS

We would like to thank the editorial team for their assistance and to the anonymous reviewers for all their input to improve this journal.

#### REFERENCES

- [1] F. N. Aini, Clustering Business Process Model Petri Net dengan Complete Linkage. ITSSMART, 2014, 3, pp. 47–51.
- [2] E. Altman, E. Hotchkiss, Corporate Financial Distress and Bankruptcy; Predict and Avoid Bankruptcy, Analyze and Invest in Distressed Debt (3rd ed.), 2006.
- [3] E. I. Altman, Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy, The Journal of Finance, 1968, Vol. 23, No.4, pp. 589–609.
- [4] Y. Andriani, D. Cahyani, V. Gusmaryanita, Discriminant Analysis to Determine Factor Affecting the Choice of Mathematics Study Programs at FMIPA and FKIP Sriwijaya University, Journal of Science Research, 2011, Vol.14, No.4, pp. 168-226
- [5] N. Gerantonis, K. Vergos, A. G. Christopoulos, Can Altman Z-score Models Predict Business Failures in Greece . Research Journal of International Studies, 2009, October(12), pp. 21–28.
- [6] I. Ghozali, Applied Multivariat Analysis IBM SPSS 23, Semarang, Diponegoro, 2016.
- [7] R. P. Jati, A. Prasetyo, Analysis of the Bankruptcy Potential of Islamic Commercial Banks in Indonesia in the 2012-2016 Period Using the Multiple Discriminant Analysis Method, Jurnal Ekonomi Syariah Teori Dan Terapan, 2019, Vol. 5(11), pp. 941.
- [8] S. Lifschutz, Predicting Bankruptcy: Evidence from Israel, International Journal of Business and Management, 2010, Vol.5(4), pp. 133–141.
- [9] U. Narimawati, Qualitative and quantitative research methodologies, theories and applications., Bandung, Agung Media, 2008.
- [10] J. Oates, W. W. Weston, J. Jordan, The impact of patient-centered care on outcomes, Fam Pract, 2000, Vol. 49(9), pp. 796–804.
- [11] H. D. Platt, M. B. Platt, Predicting corporate financial distress: Reflections on choice-based sample bias, Journal of Economics and Finance, 2002, Vol. 26(2), pp. 184–199.
- [12] A. Widarjono, Applied Multivariate Analysis (2nd ed.). Yogyakarta, UPP STIM YKPN, 2015.