http://ejournal.seaninstitute.or.id/index.php/Ekonomi
Jurnal Ekonomi, Volume 12, No 04 2023
ISSN: 2301-6280 (print) ISSN: 2721-9879 (online)

# CREDIT RATING ANALYSIS OF COAL PRODUCTION INDUSTRY USING LOGISTIC REGRESSION MODEL

**Arief Tirtana[1], Aditio Wahyudi[2], Prima Naomi[3]**
[1,2,3]Magister Management, Faculty of Economics and Business Universitas Paramadina

ARTICLE INFO

*Keywords*:
Credit Rating, Logistic Regression, Financial Distress, Coal Companies

E-mail:
arief.tirtana@students.paramadina.ac.id

ABSTRACT

Purpose: The main idea of this study is to identify the credit rating of the coal-producing sub-industry through the utilization of binary logistic regression analysis. Method: This research used 19 companies of IDX (Indonesia Stock Exchange) as samples with purposive sampling method. This paper use financial ratio data in October 2022. The variables used are six types of financial ratios, namely Price to Earning Ratio (PER), Return on Equity (ROE), Return on Asset (ROA), Net Profit Margin (NPM), Price to Book Value (PBV), and Debt to Equity Ratio (DER). This study uses binary logistic regression analysis. Result: The empirical result shown that financial ratio such as PBV, PER, and DER can be used prediction model with accuration rate of 0.95 and ROC 1.00. Based on the result, the companies with the best credit rating (AAA) is Bukit Asam Tbk (PTBA), Indika Energy Tbk (INDY), Baramulti Suksessarana Tbk (BSSR), and Adaro Energy Indonesia Tbk (ADRO). Novelty: The research presented in this study is distinguished by its utilization of a two-stage approach. The initial phase involves the application of the K-Means method to replicate the clustering of the company in order to identify binary performance data. The subsequent phases involve the utilization of binary logistic regression, incorporating six distinct financial ratios.



## 1. INTRODUCTION

The attention of both academics and practitioners has been drawn to credit risk assessment due to the ongoing advancement of commercial bank credit operations (Li & Chen, 2020). According to (Kuznyetsova et al., 2022), the inclusion of financial statements and non-financial information is a crucial aspect to be taken into account when conducting credit risk analysis for organizations. One of the strategies used for credit evaluation involves the implementation of a system that aims to assess the creditworthiness of facility grantees through credit evaluation or rating (Ershadi & Omidzadeh, 2018). The utilization of credit rating (or credit scoring) models has become widespread in the assessment of credit admission due to the significant expansion of the credit business and the administration of substantial loan portfolios. Credit rating models are designed to categorize loan customers into two groups: a good credit group (accepted) and a bad credit group (rejected). This classification is based on various characteristics, including age, income, marital status, or historical data of previously accepted and rejected applicants (Tsai & Chen, 2010). One of the advantages associated with the consideration of credit scoring is the potential reduction in the cost of credit analysis. This approach allows for speedier decision-making processes, as well as the ability to ensure credit collections and mitigate potential risks (West, 2000).

The process of determining credit ratings involves the utilization of many approaches. Historically, judgment-based procedures were widely employed for the purpose of validation. The current approach to decision-making relies on expert opinions, which are characterized by significant time consumption, high costs, subjectivity, and a lack of scientific validity. Statistical model-based methods possess a high degree of rationality, cost-effectiveness, and empirical support. Currently, the predominant focus of credit risk assessment research lies in the examination of feature selection and classifier methods (Deng et al., 2020). Previous research has indicated that certain statistical techniques, such as logistic regression, have been employed in the prediction of credit risk.

In recent years, the coal industry in Indonesia has emerged as a lucrative and appealing sector, mostly driven by the escalating global prices of coal. According to the Statistical Review of World Energy

2022, there was a significant increase in coal prices during the year 2021. The average price of coal in Europe is $121 per tonne, whereas the average price in the Asian market is $145 per tonne. The global output of goods and services aligned with consumption levels, resulting in a notable rise in supply by 440 million metric tons. The rise in production was mostly driven by China and India, with a significant portion of the output being consumed inside their respective domestic markets. Additionally, Indonesia also contributed to this growth, hence facilitating an increase in exports.

The objective of this study is to ascertain the credit rating of the coal-producing sub-industry through the utilization of binary logistic regression analysis. The present study employed a purposive selection strategy to select a sample of 19 businesses listed on the Indonesia Stock Exchange (IDX). This research utilizes financial ratio data from October 2022. The study incorporates six distinct financial ratios as variables, including the Price to Earning Ratio (PER), Return on Equity (ROE), Return on Asset (ROA), Net Profit Margin (NPM), Price to Book Value (PBV), and Debt to Equity Ratio (DER). The subsequent section in the research paper is dedicated to the research method, followed by the result and discussion section, and concludes with the final section on the conclusion.

## 2. METHOD

This study analyzes quantitative data sourced from the October 2022 edition of the Indonesia Stock Exchange (IDX) Digital Statistic (Beta). The focus of this study encompasses a sample of 19 firms that are publicly listed on the Indonesia Stock Exchange (IDX) and fall within the energy sector, namely the sub-industry of coal extraction. The utilized variables encompass six distinct financial ratios, specifically the Price to Earning Ratio (PER), Return on Equity (ROE), Return on Asset (ROA), Net Profit Margin (NPM), Price to Book Value (PBV), and Debt to Equity Ratio (DER). The utilization of financial ratios serves as a valuable tool for assessing a company's financial standing and operational performance, enabling comparisons to be made with past years or other entities (Herawati, A., & Putra, A. S, 2018). The use of financial ratios can also serve as a means to identify discrepancies in the execution of a company's operational endeavors, achieved by juxtaposing the current financial ratio with those of preceding years (Wild et al., 2007).

The methodology employed in this study consists of two distinct steps. The initial phase involves the simulation of company clustering in order to identify binary performance data using the K-Means algorithm. The subsequent steps involve the utilization of binary logistic regression, incorporating six distinct financial ratios. Next, the second binary logistic regression will be examined, which utilizes major financial ratios to assess the credit rating within the coal production industry. The software utilized in this study is Orange Data Mining.

### Clustering Method

According to (Mirkin, 1996), clustering is a mathematical method that aims to uncover classification patterns within real-world data. Clustering is a technique employed to partition a given dataset into distinct classes, accomplished through the assignment of labels to each class. This process is guided by the objective of maximizing the similarity within each class while simultaneously decreasing the similarity between different classes. Clusters are created within the dataset to facilitate the grouping of related objects together, whereas objects that exhibit significant dissimilarity are allocated to other clusters (Babu, 2012). According to (Ogbuabor & Ugwoke, 2018), clustering algorithms are employed to arrange a given data set into distinct clusters, where the data points inside a cluster exhibit similarity, while those in different clusters display dissimilarity.

This work is a demonstration of the implementation of firm clustering utilizing widely recognized clustering algorithms, specifically K-means. The K-Means algorithm is a non-hierarchical clustering technique that involves an initial determination of the number of clusters to be produced. The data is then divided into the specified number of clusters depending on the proximity of each item to the center of its respective cluster (Cebeci & Yildiz, 2015).

### Logistic Regression Method

Logistic regression is a parametric statistical method that is characterized by its simplicity. This method exhibits similarities to conventional regression analysis. Hence, it is imperative to ensure that the application of logistic regression adheres to certain assumptions of conventional regression analysis. These assumptions include the avoidance of autocorrelation in residuals, the prevention of multi-collinearity among independent variables, and the requirement for the collected data to exhibit a normal distribution (Hosmer Jr et al., 2013).

Logistic regression employs a sequence of numerical computations to construct a model based on known classification parameters, with the objective of determining the discriminatory capacity of each parameter for different groups and establishing classification rules for each group. Logistic regression and discriminant analysis are both statistical techniques employed to examine the association between independent variables and a categorical dependent variable, whereby the dependent variable consists of a set of categories for classifying objects. Hence, a notable distinction between logistic regression and discriminant analysis lies in the requirement for discriminant analysis to adhere to the assumptions of normal distribution and equal covariance matrices in order to determine the ideal value. Nevertheless, logistic regression does not require these assumptions. Furthermore, even if these assumptions are met, logistic regression can still yield reasonably accurate predictions ((Tsai & Chen, 2010).

The steps taken in this research are as follows.

1. Collecting data
   The data used was the downloaded data from IDX Digital Statistic (Beta) in October 2022.
2. Cleansing data
   There is a total of 21 enterprises classified within the sub-industry of coal production. However, there are two corporations with incomplete financial data, and one company with anomalous data. In order to mitigate bias, three companies have been omitted from the analysis due to the presence of missing values and anomalous conditions. The three aforementioned companies are Borneo Olah Sarana Sukses Tbk (BOSS), Garda Tujuh Buana Tbk (GTBO), and PT. Trada Alam Minera Tbk (TRAM). Hence, the present study incorporates a sample of 19 organizations, as depicted in Table 1, to ascertain the credit rating within the coal producing sector.

**Table 1.** Coal production industry Financial Ratio

| No | Code | Stock Name | PER | PBV | DER | ROA | ROE | NPM |
|---|---|---|---|---|---|---|---|---|
| 1 | ADMR | Adaro Minerals Indonesia Tbk | 12.20 | 9.48 | 1.54 | 0.31 | 0.78 | 0.57 |
| 2 | ADRO | Adaro Energy Indonesia Tbk | 4.41 | 1.55 | 0.60 | 0.22 | 0.35 | 0.55 |
| 3 | AIMS | Akbar Indo Makmur Stimec Tbk | 10.81 | 2.78 | 0.94 | 0.13 | 0.26 | 0.22 |
| 4 | ARII | Atlas Resources Tbk | 5.02 | 1.39 | 5.17 | 0.04 | 0.28 | 0.17 |
| 5 | BSSR | Baramulti Suksessarana Tbk | 2.76 | 2.59 | 0.37 | 0.68 | 0.94 | 0.67 |
| 6 | BUMI | Bumi Resources Tbk | 13.36 | 4.92 | 4.03 | 0.07 | 0.37 | 0.34 |
| 7 | BYAN | Bayan Resources Tbk | 7.47 | 6.55 | 0.33 | 0.66 | 0.88 | 0.64 |
| 8 | COAL | Black Diamond Resources Tbk | 14.35 | 9.18 | 1.57 | 0.25 | 0.64 | 0.28 |
| 9 | DSSA | Dian Swastatika Sentosa Tbk | 6.24 | 0.71 | 1.31 | 0.05 | 0.11 | 0.12 |
| 10 | GEMS | Golden Energy Mines Tbk | 5.62 | 5.83 | 1.11 | 0.49 | 1.04 | 0.40 |
| 11 | HRUM | Harum Energy Tbk | 7.07 | 1.70 | 0.25 | 0.19 | 0.24 | 0.55 |
| 12 | INDY | Indika Energy Tbk | 4.67 | 0.98 | 2.40 | 0.06 | 0.21 | 0.13 |
| 13 | ITMG | Indo Tambangraya Megah Tbk | 4.27 | 2.31 | 0.33 | 0.41 | 0.54 | 0.56 |
| 14 | KKGI | Resource Alam Indonesia Tbk | 5.58 | 1.95 | 0.40 | 0.25 | 0.35 | 0.35 |
| 15 | MBAP | Mitrabara Adiperdana Tbk | 3.08 | 2.78 | 0.40 | 0.64 | 0.90 | 0.58 |
| 16 | MCOL | Prima Andalan Mandiri Tbk | 4.31 | 3.06 | 0.36 | 0.52 | 0.71 | 0.52 |
| 17 | PTBA | Bukit Asam Tbk | 3.43 | 1.70 | 0.56 | 0.32 | 0.50 | 0.42 |
| 18 | SMMT | Golden Eagle Energy Tbk | 6.24 | 2.31 | 0.21 | 0.31 | 0.37 | 0.80 |
| 19 | TOBA | TBS Energi Utama Tbk | 5.77 | 0.87 | 1.23 | 0.07 | 0.15 | 0.21 |

**Source:** IDX Digital Statistic (Beta) October 2022 (2022)

1. Clustering with the K-Means algorithm
   The K-Means approach for clustering is founded on an iterative process that aims to minimize the total sum of distances between each item and its corresponding cluster centroid. The movement of objects between clusters continues until the sum can no longer be reduced (Cebeci & Yildiz, 2015).
2. Identify binary data of performance
   Based on the clustering outcome, the perform cluster and distress cluster can be defined. The two highest performing clusters have been allocated the label "1," while the cluster associated with distress has been assigned the label "0".
3. Logistic Regression
   Logistic regression employs a sequence of numerical computations to construct a model based on given classification parameters, with the aim of identifying the parameters that possess greater discriminatory capacity for each group, as well as determining the classification rules for each group. The initial experiment included a set of six financial ratios. Next, the second binary logistic regression will be examined, focusing solely on crucial financial ratios, in order to ascertain the credit rating inside the coal-producing industry.
4. Determine the Credit Rating

The credit rating can be determined using a logistic regression formula-based approach. The highest rating assigned to companies is 'AAA', denoting superior performance, while the lowest rating is 'D', indicating poor performance.

## 3. RESULT AND DISCUSSION

Prior to implementing the clustering method, it is essential to do a descriptive statistical study. Descriptive statistical analysis may also be employed to establish the pattern of data distribution. Descriptive statistics play a crucial role in summarizing raw data, enabling an understanding of the underlying distribution pattern and providing insights into the form of the data distribution. The present study employs descriptive statistical analysis to provide a comprehensive depiction of the data through the utilization of tables, graphs, and measures such as the mean, median, maximum, minimum, and variability for each variable. Based on the presented table, it is evident that the data has a normal distribution.

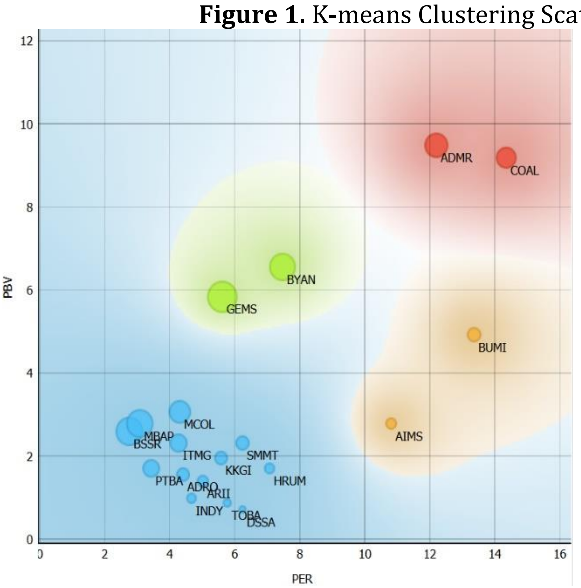
**Table 2.** Descriptive Statistics Analysis

| Financial Ratio | Min. | Max. | *Mean* | *Median* | *Dispersion* |
|---|---|---|---|---|---|
| ROE | 0.11 | 1.04 | 0.50 | 0.37 | 0.56 |
| ROA | 0.04 | 0.68 | 0.29 | 0.25 | 0.70 |
| PER | 2.76 | 14.35 | 6.66 | 5.62 | 0.50 |
| PBV | 0.71 | 9.48 | 3.29 | 2.31 | 0.78 |
| NPM | 0.12 | 0.80 | 0.42 | 0.42 | 0.45 |
| DER | 0.21 | 5.17 | 1.21 | 0.60 | 1.07 |

**Source:** IDX Digital Statistic, Oct 2022 processed with orange data mining (2022)

The initial phase in the process of K-Means clustering involves the determination of the optimal number of clusters. The K-Means technique yielded four distinct clusters, each exhibiting unique traits based on the attributes of the items within them. Figure 1 displays a scatter plot that visually represents the clustering of objects based on their proximity to the nearest centroid.

**Figure 1.** K-means Clustering Scatter Plot and the Member of each Cluster




| Code | Stock Name | Cluster | Binary |
|---|---|---|---|
| ADRO | Adaro Energy Indonesia Tbk | A | 1 |
| ARII | Atlas Resources Tbk | A | 1 |
| BSSR | Baramulti Suksessarana Tbk | A | 1 |
| DSSA | Dian Swastatika Sentosa Tbk | A | 1 |
| HRUM | Harum Energy Tbk | A | 1 |
| INDY | Indika Energy Tbk | A | 1 |
| ITMG | Indo Tambangraya Megah Tbk | A | 1 |
| KKGI | Resource Alam Indonesia Tbk | A | 1 |
| MBAP | Mitrabara Adiperdana Tbk | A | 1 |
| MCOL | Prima Andalan Mandiri Tbk | A | 1 |
| PTBA | Bukit Asam Tbk | A | 1 |
| SMMT | Golden Eagle Energy Tbk | A | 1 |
| TOBA | TBS Energi Utama Tbk | A | 1 |
| BYAN | Bayan Resources Tbk | B | 0 |
| GEMS | Golden Energy Mines Tbk | B | 0 |
| AIMS | Akbar Indo Makmur Stimec Tbk | C | 0 |
| BUMI | Bumi Resources Tbk | C | 0 |
| ADMR | Adaro Minerals Indonesia Tbk | D | 0 |
| COAL | Black Diamond Resources Tbk | D | 0 |

**Source:** data processed with orange data mining (2022)

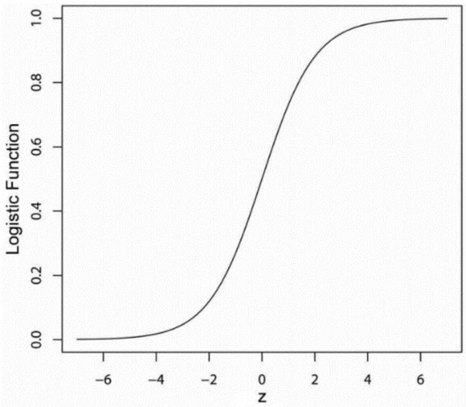
Cluster A comprises 13 enterprises, whereas clusters B, C, and D consist of 2 companies each. Cluster A is considered a performing cluster, and as a result, it is assigned the binary code 1. Clusters B, C, and D are designated with the binary code 0. The logistic regression model is distinct from other statistical methods, such as discriminant analysis or linear regression, due to its utilization of a distinct distribution function for estimation (Press & Wilson, 1978). This characteristic renders it particularly suitable for addressing credit rating concerns. Furthermore, in order to enhance the precision and adaptability of this model, several approaches have been suggested for the construction of the binary logistic regression model (Agresti et al., 1990). Indeed, it is reasonable to do logistic regression when the dependent variable consists of discrete values. The dependent variable in this study is the creditworthiness rating, which was previously described as having discrete values of zero and one. The utilization of logistic regression has gained popularity primarily as a result of the distinctive shape and behavior exhibited by the logistic function.

This model has been specifically built to articulate the likelihood, which consistently assumes values within the range of zero and one. It is prudent to employ this particular model for risk assessment due to its consistent adherence to a numerical scale ranging from zero to one, a characteristic that is not universally present in alternative models. The popularity of this model can be attributed to the presence of the S-shaped logistics function curve, as depicted in Figure 2. The logistics function curve, characterized by an S-shape, illustrates the relationship between risk and variable Z. It indicates that for small values of Z, the risk is minimal. However, once the desired threshold is surpassed, the risk begins to escalate. For instance, when Z exceeds -4 in the graph, the risk experiences a significant increase. Furthermore, when Z surpasses 4, the risk remains relatively constant. Finally, for large values of Z, the risk stabilizes around a value of one (Kleinbaum et al., 2010).

For fitting a logistic regression model, estimate the value of Z in the logistics function by $\alpha + \beta_1 X_1 + \beta_2 X_2 + \dots \beta_n X_n$, where independent variables are X1 to Xt, then with estimating the existing parameters with maximum likelihood method, we estimate the probability or dependent variable risk. In which $X_1$, ...,$X_n$ are independent variables. In addition the model can be rewritten according to odds ratio Logarithm as
follows:

$$\log \frac{p}{(1-p)} = \alpha + \beta 1\, X1 + \beta 2\, X2 + \dots . \beta n\, Xn$$

**Figure 2.** Logistic Function




**Source:** Ershadi & Omidzadeh (2018)

Finally, we used Orange Data Mining software to fit the logistic regression model to mentioned data. Where, PBV stands for Price to Book Value, PER stands for Price to Earnings Ratio, DER stands for Debt-to-Equity Ratio, NPM stands for Net Profit Margin, ROE stands for Return on Equity, and ROA stands for Return on Asset, and α is the intercept of model. Estimation of regression coefficients of the above model is specified in Table 3.

**Table 3.** Logistic Regression Result using six financial Ratio

| Financial Ratio | Correlation | Regression Coefficient |
|---|---|---|
| Intercept | | 3.04 |
| ROE | 0.09 | 1.04 |
| ROA | 0.02 | 0.63 |
| PER | 0.71 | -0.16 |
| PBV | 0.73 | -0.96 |
| NPM | 0.19 | 1.04 |
| DER | 0.24 | 0.46 |

Note. Accuracy Ratio(AR) = 0.81, Area Under Curve (AUC) = 0.75

**Source:** data processed from logistic regression result

According to calculated correlation for each variable, it can be concluded that the best three factors is PBV, PER, and DER. PBV, PER, and DER are statistically meaningful and other variables in the model can be abandoned. Therefore the regression model after removing insignificant variables is as Table 4.

**Table 4.** Logistic Regression Result using significant financial Ratio

| Financial Ratio | Regression Coefficient |
|---|---|
| Intercept | 9.76 |
| PER | -0.77 |
| PBV | -1.11 |
| DER | -1.11 |

Note. Accuracy Ratio(AR) = 0.95, Area Under Curve (AUC) = 1.00

**Source:** data processed from logistic regression result

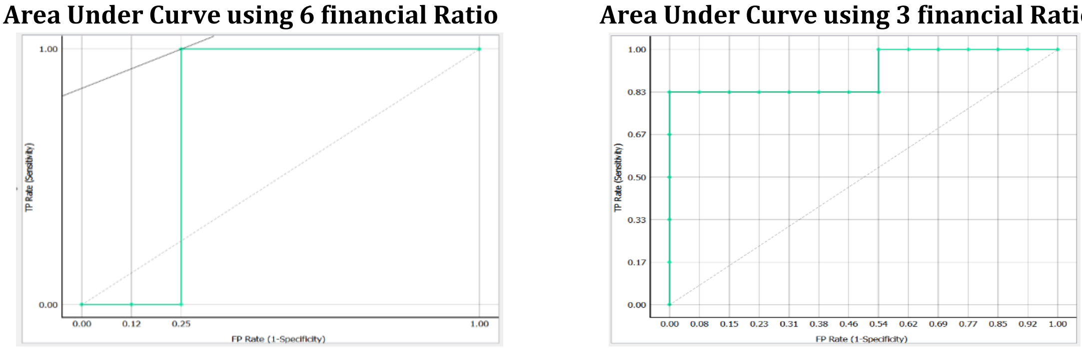
There are multiple methods available to assess the accuracy of a model. In this study, the Accuracy Ratio (AR) and Receiver Operating Characteristic (ROC) are employed to evaluate the model's correctness. The Accuracy Ratio is a metric that quantifies the balance between the selection rates of "good" and "bad" outcomes. In the event where the scoring model is characterized by randomness, it may be observed that the ratio of successful outcomes over a specific threshold will be equivalent to the ratio of unsuccessful outcomes. This would result in an annualized rate (AR) of 0%. Conversely, by attaining a flawless scorecard, it would be feasible to opt for the entirety of the desirable items (100%) while completely avoiding the undesirable ones (0%). The resultant augmented reality (AR) would achieve a perfect score of 100%. The accuracy rate (AR) achieved in this study for logistic regression utilizing a set of six financial ratios is 0.81. However, when considering only the best three financial ratios, the AR increases to 0.95.

The evaluation of credit risk models can be conducted using two distinct methodologies, namely Receiver Operating Characteristic (ROC) analysis and Cumulative Accuracy Profile (CAP) analysis. Both receiver operating characteristic (ROC) analysis and cumulative accuracy profile (CAP) analysis are effective methods for evaluating the accuracy of credit rating assessments (Irwin & Irwin, 2013). The fundamental concept that underlies the analysis of Receiver Operating Characteristic (ROC) and Cumulative Accuracy Profile (CAP) is that the process of diagnosis entails a compromise between default and non-default outcomes, specifically true and false positives. This compromise is contingent upon the level of strictness applied to the threshold used for determining when an alarm should be triggered. ROC analysis has been employed by financial analysts to evaluate credit-ratings systems and signs of financial crises (Van Gool et al., 2012).

The performance of a rating model is considered superior when the receiver operating characteristic (ROC) curve has a steeper slope at its leftmost end and when the position of the ROC curve is in close proximity to the point (0,1). It can be inferred that a higher degree of accuracy is associated with a fitting model that has a larger surface area under the curve. According to the (Supervision, 2005), a higher area under the receiver operating characteristic (ROC) curve indicates a superior performance of the model. Figure 3 displays the receiver operating characteristic (ROC) curve pertaining to small firms. The logistic regression model utilizing six financial ratios exhibits an area under the receiver operating characteristic (ROC) curve of 0.75, but the model employing the top three financial ratios achieves a perfect area under the ROC curve of 1.00. The reliability of the credit rating can be assured by placing confidence in the outcomes of models that employ major financial ratios. It is important to acknowledge that the accuracy of computations can be enhanced by lowering the number of independent variables from six to three, specifically by eliminating variables that were not statistically significant.

**Figure 3.** ROC Curve

**Area Under Curve using 6 financial Ratio** **Area Under Curve using 3 financial Ratio**




**Source:** data processed with orange data mining

PBV, PER, and DER are statistically meaningful and other variables in the model can be abandoned. Therefore the regression model after removing insignificant variables. So, the credit rating formula used in this study is as follows:

http://ejournal.seaninstitute.or.id/index.php/Ekonomi
**Jurnal Ekonomi, Volume 12, No 04 2023**
ISSN: 2301-6280 (print) ISSN: 2721-9879 (online)

$$f(x) = \frac{1}{1+e^{-x}}$$

$$f(x) = \frac{1}{1+e^{-(9.76 + (9.76+(-1.11\ PBV)+ (-0.77\ PER)+ (-1.11\ DER)))}}$$

The credit rating classification is performed by employing the formula derived from logistic regression. Moreover, based on the outcomes of the z-score analysis, a threshold is established to ascertain the financial stability of enterprises, identify those facing financial difficulty, and predict the likelihood of bankruptcy within the upcoming years. According to (Rhomadhona, 2014), the Springate model equation employs a criterion wherein a firm with a Z value less than 0.862 is classed as bankrupt, while a company with a Z value greater than 0.862 is defined as healthy. The study's findings indicate that the classification of companies can be determined based on the results of the equation. Specifically, if the Z value is less than 0.02, it is classified as a company predicted to experience bankruptcy within 2 years. If the Z value falls within the range of 0.02 to 0.5, the company is categorized as facing financial distress. On the other hand, if the Z value exceeds 0.5, the company is considered to be in a healthy state without any financial distress.

**Table 5.** The categories and distribution of credit ratings

| Code | Stock Name | Z Score | Information |
|---|---|---|---|
| PTBA | Bukit Asam Tbk | 0.9901 | No Financial Distress |
| BSSR | Baramulti Suksessarana Tbk | 0.9872 | No Financial Distress |
| ADRO | Adaro Energy Indonesia Tbk | 0.9816 | No Financial Distress |
| MBAP | Mitrabara Adiperdana Tbk | 0.9793 | No Financial Distress |
| ITMG | Indo Tambangraya Megah Tbk | 0.9719 | No Financial Distress |
| TOBA | TBS Energi Utama Tbk | 0.9519 | No Financial Distress |
| KKGI | Resource Alam Indonesia Tbk | 0.9456 | No Financial Distress |
| DSSA | Dian Swastatika Sentosa Tbk | 0.9378 | No Financial Distress |
| MCOL | Prima Andalan Mandiri Tbk | 0.9337 | No Financial Distress |
| INDY | Indika Energy Tbk | 0.9178 | No Financial Distress |
| SMMT | Golden Eagle Energy Tbk | 0.8964 | No Financial Distress |
| HRUM | Harum Energy Tbk | 0.8958 | No Financial Distress |
| ARII | Atlas Resources Tbk | 0.1999 | Financial Distress |
| GEMS | Golden Energy Mines Tbk | 0.0936 | Financial Distress |
| AIMS | Akbar Indo Makmur Stimec Tbk | 0.0634 | Financial Distress |
| BYAN | Bayan Resources Tbk | 0.0259 | Financial Distress |
| BUMI | Bumi Resources Tbk | 0.0000 | Bankruptcy predicted within 2 years |
| ADMR | Adaro Minerals Indonesia Tbk | 0.0000 | Bankruptcy predicted within 2 years |
| COAL | Black Diamond Resources Tbk | 0.0000 | Bankruptcy predicted within 2 years |

**Source:** data processed from logistic regression result

Bankrupt firms typically undergo a transitional phase characterized by a decline in financial stability, progressing from a state of possible crisis to eventual bankruptcy. The ability to anticipate imminent financial difficulty in corporations can provide valuable time for regulators, investors, and management of the distressed firm to mitigate any additional harm. Hence, it is imperative to demonstrate the timeliness with which our prediction model can effectively forecast instances of financial trouble. In order to determine a credit rating classification, the outcomes of the logistic regression are classified into the following categories. The highest credit rating designation is AAA. There exist four firms that possess a AAA credit rating, specifically Bukit Asam Tbk (PTBA), Indika Energy Tbk (INDY), Baramulti Suksessarana Tbk (BSSR), and Adaro Energy Indonesia Tbk (ADRO).

**Table 6.** The categories and distribution of credit ratings

| Z score | Rating | Percentage of Firm |
|---|---|---|
| $Z > 0.97$ | AAA | 26.32% |
| $0.94 < Z < 0.97$ | AA | 10.53% |
| $0.90 < Z < 0.94$ | A | 15.79% |
| $0.80 < Z < 0.90$ | BBB | 10.53% |
| $0.50 < Z < 0.80$ | BB | 0.00% |
| $0.10 < Z < 0.50$ | B | 5.26% |
| $0.02 < Z < 0.10$ | C | 15.79% |
| $Z < 0.02$ | D | 15.79% |

**Source:** data processed from logistic regression result

The result of Credit Rating for Coal Production Industry can be seen in Table 7.

**Table 7.** Credit Rating for Coal Production Industry

| Code | Stock Name | Z Score | Credit Rating |
|---|---|---|---|
| PTBA | Bukit Asam Tbk | 0.9901 | AAA |
| BSSR | Baramulti Suksessarana Tbk | 0.9872 | AAA |
| ADRO | Adaro Energy Indonesia Tbk | 0.9816 | AAA |
| MBAP | Mitrabara Adiperdana Tbk | 0.9793 | AAA |
| ITMG | Indo Tambangraya Megah Tbk | 0.9719 | AAA |
| TOBA | TBS Energi Utama Tbk | 0.9519 | AA |
| KKGI | Resource Alam Indonesia Tbk | 0.9456 | AA |
| DSSA | Dian Swastatika Sentosa Tbk | 0.9378 | A |
| MCOL | Prima Andalan Mandiri Tbk | 0.9337 | A |
| INDY | Indika Energy Tbk | 0.9178 | A |
| SMMT | Golden Eagle Energy Tbk | 0.8964 | BBB |
| HRUM | Harum Energy Tbk | 0.8958 | BBB |
| ARII | Atlas Resources Tbk | 0.1999 | B |
| GEMS | Golden Energy Mines Tbk | 0.0936 | C |
| AIMS | Akbar Indo Makmur Stimec Tbk | 0.0634 | C |
| BYAN | Bayan Resources Tbk | 0.0259 | C |
| BUMI | Bumi Resources Tbk | 0.0000 | D |
| ADMR | Adaro Minerals Indonesia Tbk | 0.0000 | D |
| COAL | Black Diamond Resources Tbk | 0.0000 | D |

**Source:** data processed from logistic regression result

According to the findings of this study, a significant proportion of coal production companies can be classified as financially stable entities. However, a subset of these enterprises is currently facing financial difficulties, with projections indicating a high likelihood of bankruptcy within a very short timeframe.

## 4. CONCLUSION

In conclusion, the analysis has yielded significant findings regarding the financial health of coal production companies. The application of k-means clustering identified four distinct clusters, with Cluster A designated as "1" and Clusters B, C, and D collectively assigned as "0" for binary data representation. Utilizing logistic regression with six financial ratios (PER, ROE, DER, NPM, and PBV) produced an AUC of 0.75 and an AR of 0.81. Further refinement of the logistic regression model with three crucial financial ratios (PBV, PER, DER) resulted in an AUC of 1.00 and an AR of 0.95. Notably, companies with an "AAA" credit rating were identified as Bukit Asam Tbk (PTBA), Indika Energy Tbk (INDY), Baramulti Suksessarana Tbk (BSSR), and Adaro Energy Indonesia Tbk (ADRO). This model, instrumental in recognizing potential corporate distress, serves as an early warning system for investors, analysts, and regulators. However, it is essential to acknowledge the limitations of this study. The research was conducted within a specific timeframe (October 2022), potentially affecting the generalizability of the results. Additionally, the focus solely on coal production companies may limit the broader applicability to all companies listed on the IDX. Furthermore, the reliance on logistic regression as the sole classification algorithm neglects the exploration of other potentially valuable methods. Despite these constraints, the findings offer valuable insights into corporate financial health, paving the way for future. To enhance the robustness and applicability of the research findings, it is suggested to extend the timeframe of the study. Analyzing data over multiple periods or integrating real-time data could provide a more comprehensive insight into the trends and patterns shaping the financial health of coal production companies. Additionally, widening the scope of the investigation by including companies from various industries listed on the IDX would contribute to a more holistic evaluation of the model's effectiveness. This diversified approach aims to offer a broader perspective on the practical utility of the model across different sectors. Moreover, exploring alternative classification algorithms beyond logistic regression, such as machine learning techniques or ensemble methods, holds the potential to reveal more nuanced insights and enhance the predictive accuracy of the model. A comparative analysis of results obtained from different algorithms could further bolster the credibility and reliability of the study's outcomes.

## REFERENCES

Agresti, A., Mehta, C. R., & Patel, N. R. (1990). Exact inference for contingency tables with ordered categories. *Journal of the American Statistical Association*, *85*(410), 453–458.

Babu, Ms. (2012). Clustering Approach to Stock Market Prediction. In *Int. J. Advanced Networking and Applications*.

Cebeci, Z., & Yildiz, F. (2015). Comparison of k-means and fuzzy c-means algorithms on different cluster structures. *Journal of Agricultural Informatics*, *6*(3).

Deng, Y., Wei, Y., & Li, Y. (2020). Credit Risk Evaluation Based on Data Mining and Integrated Feature Selection. *2020 IEEE International Conference on Signal Processing, Communications and Computing (ICSPCC)*, 1–4.

Ershadi, M. J., & Omidzadeh, D. (2018). Customer validation using hybrid logistic regression and credit scoring model: a case study. *Calitatea*, *19*(167), 59–62.

Hosmer Jr, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). *Applied logistic regression* (Vol. 398). John Wiley & Sons.

Irwin, R. J., & Irwin, T. C. (2013). Appraising credit ratings: does the cap fit better than the ROC? *International Journal of Finance & Economics*, *18*(4), 396–408.

Kleinbaum, D. G., Klein, M., Kleinbaum, D. G., & Klein, M. (2010). Introduction to logistic regression. *Logistic Regression: A Self-Learning Text*, 1–39.

Kuznyetsova, A., Boiarko, I., Khutorna, M., & Zhezherun, Y. (2022). Development of financial inclusion from the standpoint of ensuring financial stability. *Public and Municipal Finance*, *11*(1), 20–36.

Li, Y., & Chen, W. (2020). A comparative performance assessment of ensemble learning for credit scoring. *Mathematics*, *8*(10), 1756.

Mirkin, B. (1996). *Mathematical classification and clustering* (Vol. 11). Springer Science & Business Media.

Ogbuabor, G., & Ugwoke, F. N. (2018). Clustering algorithm for a healthcare dataset using silhouette score value. *Int. J. Comput. Sci. Inf. Technol*, *10*(2), 27–37.

Press, S. J., & Wilson, S. (1978). Choosing between logistic regression and discriminant analysis. *Journal of the American Statistical Association*, *73*(364), 699–705.

Rhomadhona, M. N. (2014). Analisis Perbandingan Kebangkrutan Model Altman, Model Springate, Dan Model Zmijewski Pada Perusahaan Yang Tergabung Dalam Grup Bakrie Yang Terdaftar Di Bursa Efek Indonesia Periode 2010-2012. *Jurnal Universitas Negeri Surabaya*.

Supervision, B. C. on B. (2005). *Studies on the validation of internal rating systems*. Bank for International Settlements Basle^ eSW SW.

Statistical Review of World Energy. (2022). bp Statistical Review of World Energy 2022 | 71st edition.

Tsai, C.-F., & Chen, M.-L. (2010). Credit rating by hybrid machine learning techniques. *Applied Soft Computing*, *10*(2), 374–380.

Van Gool, J., Verbeke, W., Sercu, P., & Baesens, B. (2012). Credit scoring for microfinance: is it worth it? *International Journal of Finance & Economics*, *17*(2), 103–123.

West, D. (2000). Neural network credit scoring models. *Computers & Operations Research*, *27*(11–12), 1131–1152.

Wild, J. J., Subramanyam, K. R., & Halsey, R. F. (2007). *Análisis de estados financieros*.