The Implementation of Logistic Regression to Develop Vendor Rating Model

Review of Integrative Business & Economics Research

Selviana
School of Business and Management ITB Jakarta

Taufik Faturohman*
School of Business and Management ITB Bandung

ABSTRACT

A vendor rating model is an essential tool to rate vendor performance for successful vendor management. This research is conducted in an Indonesian leading palm oil company facing the problem of assessing the vendors' timeliness of delivery. To address that issue, this vendor rating model is developed to help the company further assess the vendor's timeliness of delivery based on the vendor's score for risk management purposes. This scoring model is constructed by modifying the credit risk scorecard using logistic regression based on the company's historical vendors' performance data and vendors' profiles. There are 13 independent variables in this research and one dependent variable (timeliness of delivery). The results identified four variables that could be used as probability predictors of not timely delivery: age of parent company, digital existence, type of company, and long-term contract agreement existence. A cross-validation test is used to test the model, with the result as follows: the optimum fitted cut-off value is 0.53 with the accuracy rate of 100%, sensitivity rate of 70%, and misclassification rate of 17.7%. This is the first research to develop a vendor rating model using credit risk scorecard and logistic regression to the researchers' best knowledge.

Keywords: vendor rating model, vendor evaluation, risk management, logistic regression

1.INTRODUCTION

1.1 Background

Angelise Agri is a wholly-owned subsidiary of Angara Agri, with headquarters in Singapore. Angara Agri is a public company, one of the leading integrated palm-based consumer companies in the world that focuses on sustainable palm oil production. Angelise Agri's essential activities begin from cultivating its palm oil plantations in Indonesia, which includes plasma smallholders; harvesting and extracting fresh fruit bunches (FFB) into crude palm oil (CPO) and palm kernel (PK); to converting it into an extensive variety of industrial and consumer products, including cooking oil, margarine, shortening, biodiesel, and oleo-chemicals; as well as retailing and distributing palm products around the globe.

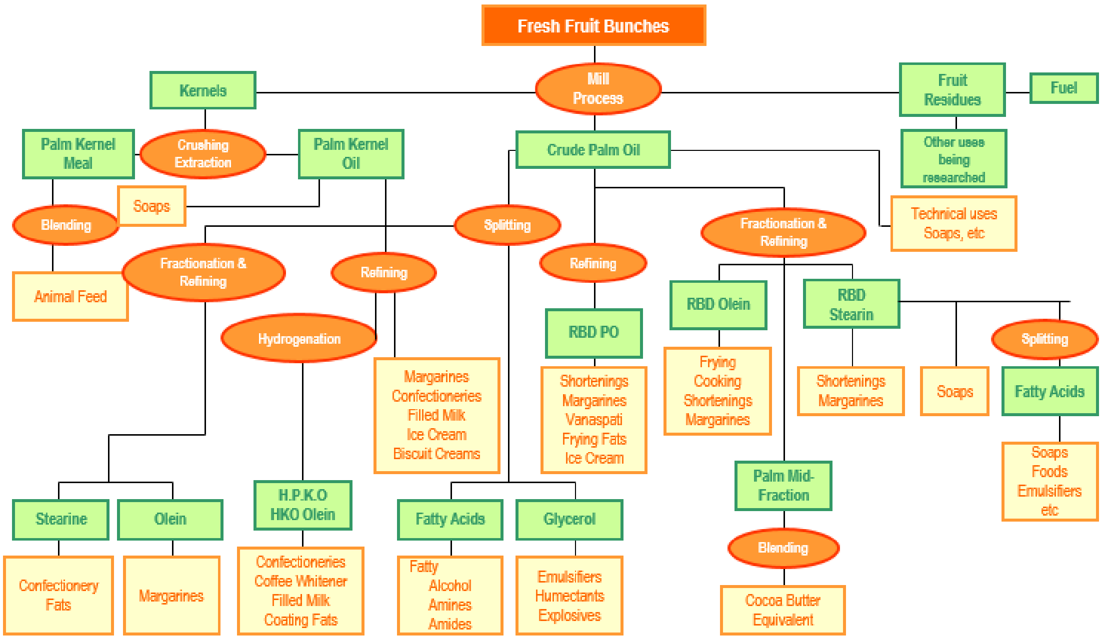

Figure 1. Palm Oil Utilization Chart

The CPO is then processed further into various brand, industrial, and value-added bulk products through its own refineries. The PK is crushed in its kernel crushing plants that produce more valuable palm kernel oil (PKO) and palm kernel meal (PKM). Angelise Agri also exports palm-based consumer products. In addition to industrial oils and bulk oils, refined products are also sold under various brands, which are known for their high quality and hold important market shares in their respective market segments in Indonesia.

In procuring its raw material, Angelise Agri's own palm oil plantation cannot fulfill its high demand; thus, Angelise Agri also procures the raw material (CPO, PK, and PKO) from third-party suppliers. In 2020, Angelise Agri procured the raw material from more than 300 vendors. The problem arose as 95% of the vendors required Angelise Agri to pay in advance prior to shipment, while Angelise Agri does not have a vendor monitoring tool to assess vendors' worthiness. In the past, Angelise Agri had experiences of giving prepayment to several vendors and ended up the goods were delivered not in a timely basis; furthermore, some could not fulfill the agreement and went default. On a weekly basis, the Commercial Controller of Angelise Agri releases a report to the management team about the update regarding the outstanding prepayment that has been paid to the vendors with the percentage of overdue (not timely) delivery, as depicted in Figure 2.

The figure only provides Angelise Agri's management team with the outstanding historical prepayment with the overdue delivery trend with no prevention tool to avoid overdue delivery from vendors. As shown on the chart, it is concluded that the overdue delivery portion remains flat throughout the year. The overdue delivery portion is, on average, 30% of the total outstanding advance payment. To address that issue, the researchers developed a vendor rating model to help Angelise Agri further assess the vendors' timeliness of delivery based on the vendor's score for risk management purposes.

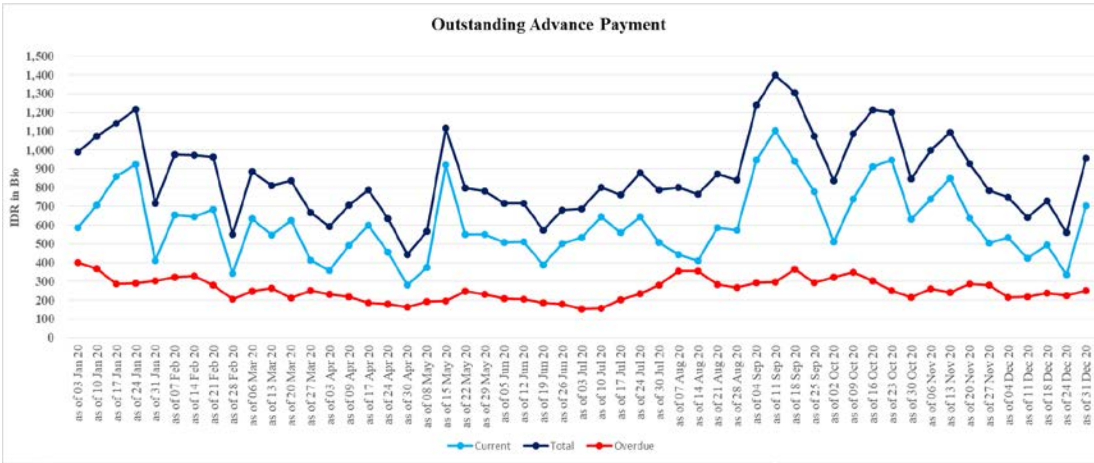

Figure 2. Outstanding Advance Payment of Angelise Agri in 2020

1.2 Objective

According to the problem statement in the previous section, this research aims to assess the predictability level of vendor rating by using historical transactions that happened in 2020 and the vendors' profiles as predictors.

2. LITERATURE REVIEW

2.1 Credit Risk

As stated by Basel (2000), credit risk is interpreted as the possibility that a counterparty will fail to meet its obligation in accordance with the agreed terms. In this research, credit risk occurs when Angelise Agri paid the advance payment prior to receiving the raw materials, as there is a possibility that the vendor may fail to deliver the raw materials on a timely basis as agreed. Assessing credit risk is crucial for companies because of the high risks associated with poor credit decisions that can lead to huge losses. This is a significant challenge today because financial institutions have faced major challenges and competition in the past decade (Lee and Chen, 2005). Therefore, this research will focus on developing a vendor rating model based on the credit risk scorecard model as the credit risk scorecard model is one of the primary methods for evaluating credit risk.

2.2 Credit Scoring Model

Credit scoring evaluates lending risk to an organization or an individual (Paleologo, Elisseeff, and Antonini, 2010). Credit models can be used in many practical applications, especially for banks and financial institutions. For example, the decision-making process to accept or reject a loan is evaluated by the bank through credit scoring so that the loan applicant can be evaluated as "good credit" – the one that is likely to be able to pay off financial liabilities or " poor credit" – the one with a high probability of default on financial liabilities (Yap, Ong, and Husain, 2011). Based on Khilfah and Faturohman (2020), the credit scoring model's advantage is that it is more accurate and transparent than other rating systems that still depend on judgment. Even though the credit scoring method is widely used by financial and banking institutions for loan applications, in this research, the researchers use it for predicting late delivery.

According to Wiginton (1980), there has been considerable interest in using quantitative models of consumer behavior for credit-granting decisions, and the logistic regression framework has been the most used statistical method for a long time. However, a more pragmatic approach has recently been carried out, from machine learning, data mining, and artificial intelligence. The researcher finds the study of machine learning approach for credit scoring in Ampountolas et al. (2021), data mining and decision tree method in Yap, Ong, and Husain (2011), and a two-stage hybrid credit scoring model using artificial neural networks and multivariate adaptive regression splines in Lee and Chen (2005). Wang et al., in their research in 2020, focused on the comparative evaluation of the performances of five popular classifiers involved in machine learning used for credit scoring: Naive Bayesian Model, Logistic Regression Analysis, Random Forest, Decision Tree, and K-Nearest Neighbor Classifier. Wang et al. (2020) stated that each classifier possesses its strength and weakness, and it is assertive to say which one is the best.

Dong, Lai, and Yen (2010) stated that credit risk scorecards had been built using various credit scoring methodologies. Among them, the logistic regression model is the most widely employed because of its excellent characteristics (e.g., robustness and transparency). Therefore, in this research, the researcher addresses the credit risk scorecard with logistic regression.

2.3 Vendor Performance Evaluation

The company's performance is increasingly dependent on the performance of external vendor partners, so effective management of these extended supply chains requires companies to adopt strategies to measure and improve the performance of network participants (AberdeenGroup, 2002).

Vendor performance evaluation is defined as the process of assessing, measuring, and observing vendor performance and vendor's business process and practicing to reduce costs, mitigate risks, and driving continuous improvement in value and operations (Gordon, 2008 p. 4). In addition, the use of a vendor performance evaluation process can improve the strength of the relationship between vendors' process innovativeness and the buyer's performance benefits (Azadegan, 2011).

In one industry benchmark report in 2002, AberdeenGroup reported that 70% of the companies viewed vendor performance evaluation as "critical" to their overall operations, yet only about half have established formal procedures for evaluating vendor performance. Thus, failure to evaluate the vendors' performance will lead to higher company costs, damage its product quality, and hinder its competitiveness in the market.

According to Gordon (2008), vendor evaluation is often more subjective; thus, in this research, the researcher aims to evaluate vendors objectively by using a quantitative approach. There is no robust methodology to develop a vendor rating model without subjectivity (personal judgment) based on researchers' best knowledge. Therefore this research aims to develop a vendor rating model objectively by modifying the credit risk scorecard.

2.4 Weight of Evidence

The predictive power of an independent variable concerning the dependent variable is determined by the Weight of Evidence (Bhalla, 2015). Credit scoring is commonly defined as a measure of the separation of good and bad customers. Customers that have defaulted on a loan are referred to as "Bad Customers." Customers that have paid back their loans are referred to as "Good Customers." According to SAS Institute, Inc, (2009), the logarithm of the ratio of the proportion of "goods" in the attribute over the proportion of "bads" in the attribute is the attribute's Weight of Evidence. High negative numbers indicate a high risk, whereas high positive values indicate a low risk. The calculation of an attribute Weight of Evidence is as follows:

$$\text{Weight of Evidence}_{\text{attribute}} = \log \frac{P_{\text{good attribute}}}{P_{\text{bad attribute}}}$$

Where:

$$P_{\text{good attribute}} = \frac{\#\,\text{goods}_{\text{attribute}}}{\#\text{goods}}$$

$$P_{\text{bad attribute}} = \frac{\#\,\text{bads}_{\text{attribute}}}{\#\text{bads}}$$

2.5 Information Value

According to Bhalla (2015), one of the most critical techniques for selecting essential variables in a predictive model is Information Value. It facilitates the classification of variables according to their significance. The Information Value is determined by applying the following formula:

$$\text{Information Value} = \sum \left(P_{\text{good attribute}} - P_{\text{bad attribute}}\right) \times \text{Weight of Evidence}$$

Based on Yap, Ong, and Husain (2011), the Information Value assessment was used to examine the ability to distinguish between high and low risks. This makes it simpler to choose which variables to include in the credit scoring model. The weighted sum of the weights of evidence of the variable's attributes is the Information Value. The weights represent the difference between the proportions of ‘goods' and ‘bads' in each attribute.

For a characteristic to be considered for inclusion in the scorecard, its Information Value must be greater than 0.02. Information values of less than 0.1 are regarded weak, those of less than 0.3 are considered medium, and those of less than 0.5 are considered strong. If the Information Value is greater than 0.5, the characteristic is overpredicting, which means it is trivially related to the good/bad information in some manner (SAS Institute, Inc, 2009).

2.6 Logistic Regression

According to Swaminathan (2018), the biological sciences employed logistic regression in the early twentieth century. It went on to be applied in a variety of social science applications. By fitting data to a logistic curve, logistic regression, also known as the logistic model or logit model, evaluates the relationship between multiple independent variables and a categorical dependent variable and calculates the probability of occurrence of an event. Binary logistic regression and multinomial logistic regression are the two types of logistic regression models. When the dependent variable is dichotomous, and the independent variables are either continuous or categorical, binary logistic regression is commonly employed. Multinomial logistic regression is used when the dependent variable is not binary and has more than two categories (Park, 2013). The purpose of logistic regression in credit risk scorecard is to calculate the conditional probability of a particular applicant belonging to a class (defaulter or non-defaulter) depending on the values of the credit applicant's independent variables (Yap, Ong, and Husain, 2011).

A logistic regression analysis indicates how the characteristics should be weighed against one another (SAS Institute, Inc, 2009). The probability of a dichotomous result (Y = 1) is related to a set of potential predictor variables in logistic regression, a commonly used statistical modeling approach (Yap, Ong, and Husain, 2011). The logistic regression model is written in the following manner:

$$\log\left(\frac{p}{1-p}\right) = w_0 + \sum w_i \log x_i$$

Where, x_i is the independent variable selected to be included in the vendor rating model.

3. RESEARCH METHODOLOGY

This research focused on evaluating the impact of historical transactions and vendors' profiles to assess the predictability level of vendors' timeliness of delivery to Angelise Agri. The purpose is to help Angelise Agri to minimize its risk of losing money by giving prepayment to the vendors with not timely delivery. The method would be using a credit scorecard and logistic regression model. This section will identify and define the steps that must be taken to answer the research objective.

The vendor rating model proposed to the company is constructed based on Angelise Agri's purchase in 2020. Angelise Agri purchased the raw materials from 180 vendor groups for that year, with more than 300 vendor entities. In this research, the top 30 vendor groups with the highest transaction volume are selected. Those top 30 vendors already represent 67% of the total volume purchased by Angelise Agri in 2020. The data used is the vendors' profile and history of transactions. Data then being processed in SPSS statistical program to build the model.

3.1 Conceptual Framework

The conceptual framework is one of the main essential aspects of the research process. It is how the research problem will best be explored, the specific direction that research should take, and the relationship between various variables in the research (Grant & Osanloo, 2014). The conceptual framework of this research is depicted in Figure 3.

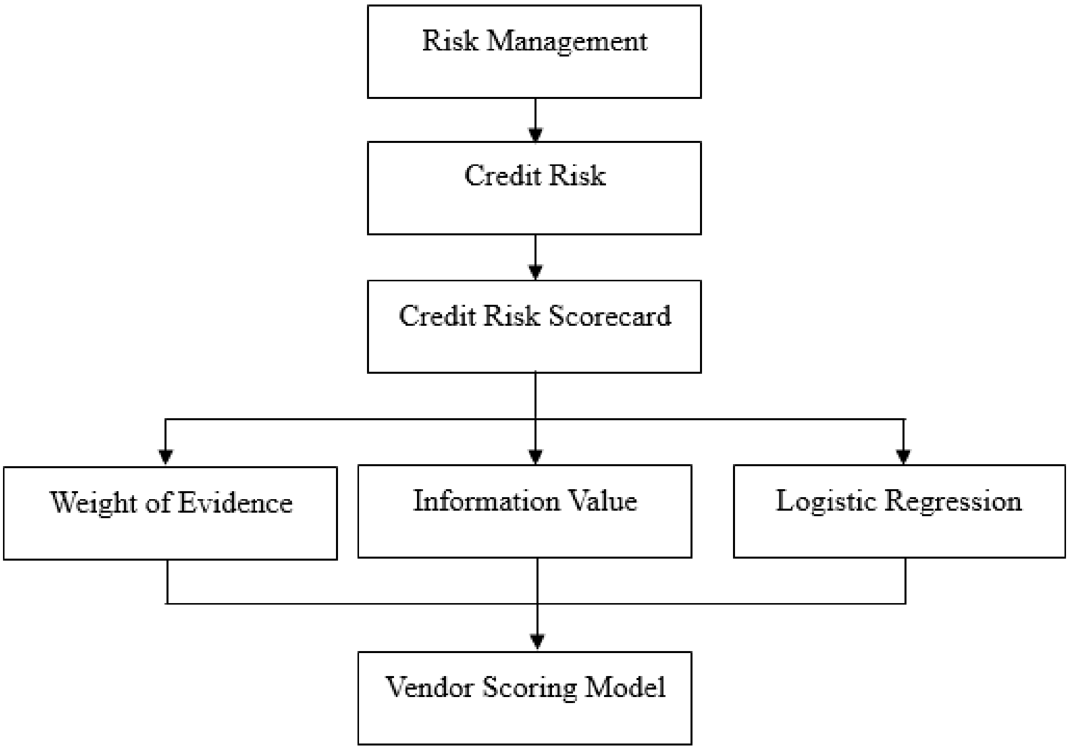

Figure 3. Conceptual Framework

According to Figure 3 above, the conceptual framework begins with a business issue: risk management of the advance payment. The business issue is then solved by developing a credit risk scorecard model, a credit risk management tool. Next, the credit risk scorecard is established by analyzing the Weight of Evidence, Information Value, and Logistic Regression. Finally, the result of Weight of Evidence, Information Value, and Logistic Regression will construct the vendor rating model.

3.2 Data Collection

The primary data is acquired by the researchers solely for the purpose of this research. In order to achieve the objectives, the researchers collected the data from the Commercial Controller of Angelise Agri. The collected data from the company are the history of transaction and length of the relationship, while the vendors' profile is investigated manually by the researchers through several websites. The researchers obtained from literature consist of journals, books, online newspapers, and websites for secondary data.

3.3 Data Processing and Analysis Methods

SPSS and Microsoft Excel are used to analyze the data. This research necessitates the following analysis:

3.3.1 Normality Test

According to Gupta et al. (2019), the most important continuous probability distribution, the standard normal distribution, has a bell-shaped density curve represented by its mean and standard deviation, and extreme values in the data set have no impact on the mean value. Because normal data is an underlying assumption in parametric testing, determining the normality of data is a need for many statistical tests. Normality can be assessed in two ways: graphically and numerically (including statistical tests). The Shapiro–Wilk test, Kolmogorov–Smirnov test, skewness, kurtosis, histogram, box plot, P–P Plot, Q–Q Plot, and mean with standard deviation are some of the most used methods for testing the normality of continuous data.

This research will address the normality test through P-P Plot. A P–P plot (also known as a probability-probability plot or percent–percent plot) is a graphical tool for determining how closely two data sets (observed and predicted) coincide. When data is normally distributed, it creates an approximate straight line. Deviations from this straight line represent deviations from normality (Ghasemi and Zahediasl, 2012).

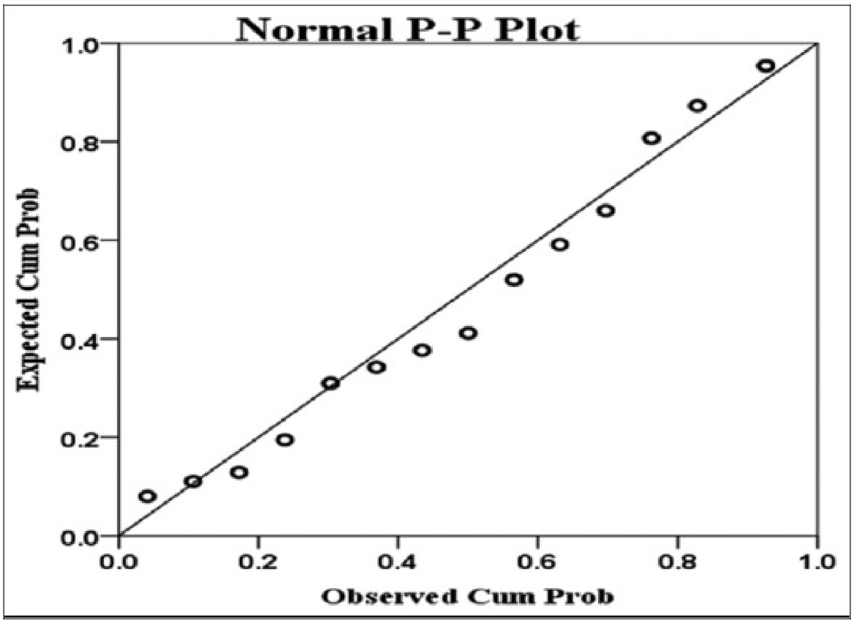

Figure 4. Normal P-P Plot
Source: Gupta et al. (2019)

3.3.2 Multicollinearity Test

According to Veronica and Anantadjaya (2014), a multicollinearity test is required to see the potential similarities with other independent variables in the model. According to Frost (2017), when independent variables in a regression model are correlated, this is called multicollinearity. Because independent variables should be independent, the correlation is an issue. When it comes to fitting the model and interpreting the results, a high degree of correlation between variables might cause issues. In a regression model, multicollinearity might be an issue since we will not be able to distinguish between the independent variables' individual impacts on the dependent variable (Bhandari, 2020). As stated by Midi, Sarkar, and Rana (2010), the test of tolerance and the Variance Inflation Factor (VIF) can be used to detect multicollinearity. According to the rule of thumb, if the

VIF of a variable exceeds 10, it has a higher collinearity level, and if the tolerance level is closer to zero, it has a higher degree of collinearity with other variables.

3.3.3 Scoring Calculation

According to Siddiqi (2006), the first step to compute the score is to calculate the scorecard points, as follows:

$$\text{Score} = \log(\text{odds}) \times \text{factor} + \text{offset}$$

$$\text{Factor} = \frac{\text{points to double odds}}{\log 2}$$

$$\text{Offset} = \text{score} - \text{factor} \times \log(\text{odds})$$

After that, the scoring calculation is computed as follows:

$$-\left(\text{WOE}_i \times \beta_i + \frac{\alpha}{n} \times \text{factor} + \frac{\text{offset}}{n}\right)$$

Where,

WOE = Weight of Evidence for each grouped attribute
β = Regression coefficient for each characteristic
α = Intercept term from logistic regression
n = Number of characteristics

3.4 Predicting Model Goodness of Fit

3.4.1 Cross-validation Test

The results of the model must be validated once the final scorecard has been selected. Validation is done to check that the solutions obtained are applicable to the subject population and have not been overfitted. According to Siddiqi (2006), it is advised that the modeling be done with a random 70% or 80% of the development sample, with the remaining 30% or 20% being "holdout sample" be kept for validation. However, if the scorecard is being created on a small sample, it may be essential to validate it on a number of randomly selected 50% to 80% samples. As the scorecard that is being developed in this research is considered minor, the researchers will proceed with randomly selected 50% samples.

3.4.2 Receiver Operating Characteristic Curve (ROC Curve)

The Receiver Operating Characteristic (ROC) curve visually depicts sensitivity (percentage of defaulters accurately predicted as defaulters) versus 1-specificity (percentage of non-defaulters inaccurately categorized as defaulters), or the true positive rate versus false-positive rate ratio (Yap, Ong, and Husain, 2011). According to Hajian-Tilaki (2013), ROC analysis is used to assess how systems can precisely distinguish two different states, usually called "default" and "non-default" ROC analysis. A ROC curve is built on the concept of a "separator" scale, on which the default and non-default distributions constitute a pair of overlapping distributions. The slope of a ROC curve at every point is equal to the likelihood ratio, which is the ratio of the two density functions representing the distribution of the separator variable in the default and non-default. A concave ROC curve relates to a monotonically increasing likelihood ratio. Instead of being dependent on a specific operating point, the Area Under the Curve (AUC) describes the whole location of the ROC curve. The AUC is a valuable measure of sensitivity and specificity that indicates the intrinsic validity of diagnostic tests. The area under the ROC curve is between 0.5 and 1.0, with higher values indicating a better model (Hajian-Tilaki, 2013).

3.5 Data Classification

There are thirteen variables as independent variables and one variable as a dependent variable. The impact of independent variables on the dependent variable will be assessed in this research. The description of each variable is shown in Table 1.

Variable Name	Description of Variable	Role
Timeliness of Delivery	The output of vendors' timeliness of delivery (timely or not timely)	Dependent
Transaction Value	Value of transaction in 2020	Independent
Transaction Volume	The volume of transactions in 2020	Independent
Transaction Frequency	Total transaction occurrence in 2020	Independent
Transaction Existence per Month	Existence of transactions in each month in 2020	Independent
Age of Parent Company	Number of the age of parent company in years	Independent
Age of the Company	Number of the age of the company in years	Independent
Digital Existence	Company website ownership	Independent
Vendor Relationship Length of Time	Number of years of transaction	Independent
Type of Company	Existence of the company in the stock exchange	Independent
Long Term Contract Agreement Existence	Existence of long-term contract agreement in 2020	Independent
Parent Company RSPO Certification Ownership	RSPO certification ownership of parent company	Independent
Company RSPO Certification Ownership	RSPO certification ownership of the company	Independent
Foreign Ownership	Foreign ownership of the company	Independent

Table 1. List of Variables

3.6 Model Development

The model is developed by modifying the credit risk scorecard; this model is derived from Yap, Ong, and Husain (2011) because by considering the predictive power and risk for each independent variable, it can focus directly on the most critical variables. The logistic regression equation is as follows:

There are 13 variables used in this research, 6 of the variables are related to the transaction history, which are: transaction value (Value), transaction volume (Volume), transaction frequency (Frequency), transaction existence per month (Routine), vendor relationship length of time (Duration), and long-term-contract agreement existence (LTC). Seven of the variables are related to the vendor's profile, which are: age of parent company (Age_P), age of the company (Age_C), digital existence (Web), type of company (Type), parent company RSPO certification ownership (RSPO_P), company

RSPO certification ownership (RSPO_C), and foreign ownership of the company (Foreign). The types of measurement used are categorical and ratio discrete. A categorical variable assigns each unit of observation into a particular group.

3.7. Data Processing

The collected data is then processed and classified by using Microsoft Excel and SPSS software. The first step is checking the Normality and Multicollinearity of the observations. The second step is calculating the Weight of Evidence of each category of the independent variable. Higher negative values represent higher risk, and higher positives values represent lower risk. In this research, the coding sequence is based on higher to lower risk. The third step of data processing is calculating the Information Value. Last but not least, the fourth step is performing the Logistic Regression in SPSS. The goodness of fit of each data processing stage is as follows:

- ✓ Normality and Multicollinearity Test
  When the data is normally distributed, the Collinearity statistics tolerance score is higher than 0.1 and VIF is less than 10; it is entitled to enter the model.
- ✓ Weight of Evidence
  When the variable has a complete logical trend, it is entitled to enter the model.
- ✓ Information Value
  When the information value is more significant than 0.02, it is to be considered to enter the model (Yap, Ong, and Husain, 2011).
- ✓ Logistic Regression
  When the variable has a probability value of less than 0.05, it is considered significant and entitled to enter the model.

4. RESULTS AND ANALYSIS

4.1 Results

4.1.1 Assumption Test (Normality and Multicollinearity)

As the amount of observation of this research is 100 data, it is assumed normal, and we may be able to relax the normality assumption (Gujarati, 2004). According to Ghasemi and Zahediasl (2012), the normality of the data can also be assessed visually, although by having a visual inspection of the distribution, readers can judge the distribution assumption by themselves. The visual method that is used in this research is the P-P plot. The P-P plot plots the cumulative probability of a variable against the cumulative probability of normal distribution. If the data are normally distributed, the result would be a straight diagonal line. The distribution of this research data is depicted in Figure 5 below:

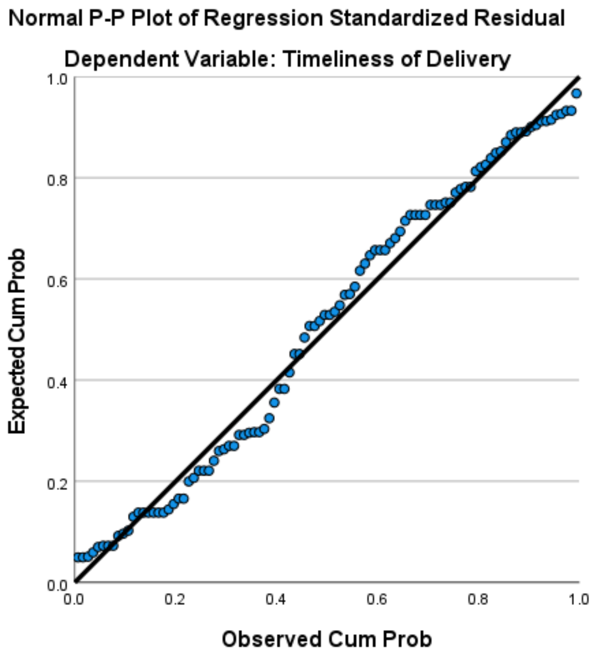

Figure 5. P-P plot of the data

As depicted in Figure 5, the P-P plot also supports the assumption that the data is normally distributed because it forms a straight diagonal line.

The next step is assessing multicollinearity. A multicollinearity test is performed to identify if correlation happens between independent variables. The result of multicollinearity is shown in Table 2 below:

Model	**Standardized Coefficients**	**T**	**Sig.**	**Collinearity Statistics**	
	Beta			**Tolerance**	**VIF**
(Constant)		6.305	0.000		
Value	-0.171	-0.808	0.422	0.155	6.453
Volume	0.057	0.271	0.787	0.158	6.331
Frequency	-0.060	-0.485	0.629	0.453	2.210
Routine	-0.074	-0.643	0.522	0.524	1.908
Age_P	-0.426	-3.769	0.000	0.546	1.831
Age_C	-0.053	-0.576	0.566	0.820	1.220
Web	-0.234	-1.810	0.074	0.416	2.402
Duration	-0.088	-0.978	0.331	0.865	1.156
Type	-0.463	-2.520	0.014	0.206	4.844
LTC	-0.161	-1.413	0.161	0.538	1.857
RSPO_P	-0.049	-0.284	0.777	0.235	4.258
RSPO_C	-0.106	-0.950	0.345	0.558	1.794
Foreign	-0.016	-0.141	0.888	0.539	1.854

Table 2. Multicollinearity Test

As depicted from the above table, all of the variables have a tolerance score of more than 0.1 and VIF < 10. They conclude that no multicollinearity exists or each independent variable has no correlation, indicating that all independent variables are entitled to enter the model.

4.1.1 Weight of Evidence

The result of the Weight of Evidence calculation is shown in Table 3 below:

Variable Name	Coding	Category	WoE	Decision
Value	1	< 100 bio IDR	-0.520	Input
	2	≥ 100 bio IDR	0.265	
Volume	1	< 12,000 MT	-0.551	Input
	2	≥ 12,000 MT	0.256	
Frequency	1	< 24	-0.915	Input
	2	24 - 48	-0.328	
	3	> 48	0.861	
Routine	1	Non-routine	-0.539	Input
	2	Routine	0.442	
Age_P	1	< 25 years	-0.798	Input
	2	≥ 25 years	0.710	
Age_C	1	< 25 years	-0.167	Input
	2	≥ 25 years	0.653	
Web	1	No	-0.091	Input
	2	Yes	0.058	
Duration	1	< 3 years	-0.733	Input
	2	≥ 3 years	0.156	
Type	1	Public	-0.503	Input
	2	Private	0.395	
LTC	1	Yes	-0.370	Input
	2	No	0.459	
RSPO_P	1	Yes	-0.486	Input
	2	No	0.337	
RSPO_C	1	No	-0.088	Input
	2	Yes	0.471	
Foreign	1	Yes	-0.145	Input
	2	No	0.034	

Table 3. Weight of Evidence

From the Weight of Evidence result above, all variables have a complete logical trend. Thus, all variables pass the goodness of fit criteria for Weight of Evidence.

4.1.2 Information Value

The result of the IV calculation is shown in Table 4 to determine the predictive power for each variable:

Variable Name	Information Value	Ordering	Predictive Power	Decision
Frequency	0.603	1	Over-predicting	Input
Age_P	0.541	2	Over-predicting	Input
Routine	0.234	3	Medium	Input
Type	0.195	4	Medium	Input
LTC	0.168	5	Medium	Input
RSPO _P	0.162	6	Medium	Input
Volume	0.140	7	Medium	Input
Value	0.136	8	Medium	Input
Duration	0.113	9	Medium	Input
Age_C	0.108	10	Medium	Input
RSPO_C	0.041	11	Weak	Input
Web	0.024	12	Weak	Input
Foreign	0.005	13	Unpredictive	Reject

Table 4. Information Value

In reference to Table 4, the result stated that two variables (Company RSPO Certification Ownership and Digital Existence) have weak predictive power. However, those variables are still included in the model because the WOE showed the logical trend order, also supported by Yap, Ong, and Husain (2011) that when Information Value is more significant than 0.02, it is to be considered to enter the model. Two variables with over-predicting power or IV more than 0.5: Transaction Frequency and Age of Parent Company, are still included as the input because their Weight of Evidence showed the logical trend order. The over-predicting power variables were still included in the model constructed by Franata et al. (2018). Therefore, there are 12 variables to enter the model as the input for IV decision which are ordered from highest to the lowest predictive power, respectively, are Transaction Frequency, Age of Parent Company, Transaction Existence per Month, Type of Company, Long Term Contract Existence, Parent Company RSPO Certification Ownership, Transaction Volume, Transaction Value, Vendor Relationship Length of Time, Age of the Company, Company RSPO Certification Ownership, and Digital Existence.

4.1.3 Initial Logistic Regression

Firstly, Logistic Regression is performed for the whole 13 independent variables to determine the significance level of each variable to be included as consideration for the goodness of fit. Yap, Ong, and Husain (2011) stated that in developing a scorecard, the discrete variables must be categorized in step with the logical trend from the WOE result. The output of initial logistic regression is stipulated as follows:

	Coef.	Std. Error	Sig.
Value	-2.043	1.462	0.162
Volume	1.038	1.463	0.478
Frequency	-0.153	0.458	0.739
Routine	-0.635	0.706	0.369
Age_P	-4.078	1.242	0.001
Age_C	-0.523	0.920	0.570
Web	-3.144	1.359	0.021
Duration	-0.816	0.756	0.281
Type	-4.870	1.604	0.002
LTC	-1.798	0.909	0.048
RSPO_P	-0.417	1.143	0.715
RSPO_C	-0.317	0.886	0.721
Foreign	-0.304	0.883	0.731
Constant	27.850	6.929	<0.001

Table 5. Initial Logistic Regression

According to Table 5, there are only four out of thirteen variables that have a significance level less than 0.05 (p-value < 0.05), which are Age of Parent Company (0.001), Digital Existence (0.021), Type of Company (0.002), and Long Term Contract Agreement Existence (0.048).

4.1.4 The Goodness of Fit

The next step is to assess the Goodness of Fit of each independent variable by looking at each WOE, IV, and P-value to determine if each variable should be inputted in the model. The result is as follows:

Variable	WoE	Information Value	P-value	Decision
Value	Yes	Yes	No	No
Volume	Yes	Yes	No	No
Frequency	Yes	Yes	No	No
Routine	Yes	Yes	No	No
Age_P	Yes	Yes	Yes	Yes
Age_C	Yes	Yes	No	No
Web	Yes	Yes	Yes	Yes
Duration	Yes	Yes	No	No
Type	Yes	Yes	Yes	Yes
LTC	Yes	Yes	Yes	Yes
RSPO_P	Yes	Yes	No	No
RSPO_C	Yes	Yes	No	No
Foreign	Yes	No	No	No

Table 6. Goodness of Fit

Based on Table 6 above, four variables (Age of Parent Company, Digital Existence, Type of Company, and Long Term Contract Agreement Existence) fulfill WOE, IV, and P-value; thus, those four variables will be included in the model.

4.1.5 Final Logistic Regression

Finally, the final logistic regression model is developed by using the selected independent variables. The result is as follows:

	Coef.	Std. Error	Sig.
Age_P	-3.904	1.134	0.001
Web	-2.213	1.013	0.029
Type	-4.547	1.388	0.001
LTC	-1.231	0.704	0.080
Constant	18.151	4.932	<0.000

Table 7. Final Logistic Regression

Based on Table 7 above, Age of the Parent Company, Digital Existence, and Type of Company is significant at a 95% confidence level, while Long Term Contract Agreement Existence is significant at a 90% confidence level. Based on the information in the table above, the equation is as follows:

$$\log\left(\frac{p}{1-p}\right) = 18.151 - 3.904\,(A) - 2.213\,(B) - 4.547\,(C) - 1.231\,(D)$$

Where,
A: the unit changes in the Age of Parent Company variable
B: the unit changes in the Digital Existence variable
C: the unit changes in the Type of Company variable
D: the unit changes in the Long-Term Contract Agreement Existence variable

The coefficients and constant are later utilized as a part of the scoring calculation process. The negative value of each coefficient demonstrates an inverse relationship between the independent variables and the dependent variable. A higher negative coefficient implies that the likelihood of being late is lower. From the above equation, it can be deciphered as follows:

- Every unit change in the Age of Parent Company variable will decrease the log of not timely delivery compared to timely delivery by 3.904 times.
- Every unit change in the Digital Existence variable will decrease the log of not timely delivery compared to timely delivery for 2.213 times.
- Every unit change in the Type of Company variable will decrease the log of not timely delivery compared to timely delivery for 4.547 times.
- In every unit change in the Long-Term Contract Agreement Existence variable will decrease the log of not timely delivery compared to timely delivery for 1.231 times.

The following statistical process is to perform logistic regression on all categories of each variable. The independent variables used are the same as the previous model, and the step-by-step process is also the same as the previous process. However, the researchers set the categorical on and consider the first category (the lowest code) as the reference category because it has the highest risk of all categories. The logistic regression results for each category are as follows:

		Coef.	Std. Error	Sig.
Age_P	≥ 25 years	-3.904	1.134	<0.001
Web	Yes	-2.213	1.013	0.029
Type	Private	-4.547	1.388	0.001
LTC	Yes	-1.231	0.704	0.080
Constant		6.256	1.701	<0.001

Table 8. Logistic Regression Model with Category

4.2 Scoring Calculation

The first step to calculate the score for each category is to determine the scaling format. The probability of not timely delivery is considered to be 1:1 due to total vendors with timely delivery : total vendors with not timely delivery are 51:49; then it is rounded to 1:1. Therefore, the odds of the scoring calculation is 1. The odds ratio compares whether the probability of a particular event is the same for two groups (Szumilas, 2010). An odds ratio of 1 indicates that the event is equally likely in the two groups. The baseline score is set to 500, and point to double odds (pdo) is set as 20 (Siddiqi, 2006) because these are the commonly used baseline score and point to double odds. The next step is to calculate the factor and offset value. The result is as follows:

Parameter	Value
Odds	1
Pdo	20
Score	500
Factor	$\frac{20}{\ln(2)} = 28.8539$
Offset	500 – (28.8539 x ln(1)) = 500

Table 9. Parameter Value

After identifying the scoring calculation parameter, the score for each category is calculated. As a result, the following is the outcome:

Variable	Coding	Group Category	Weight of Evidence	Coef.	Score
Age_P	1	< 25 years	-0.798	-3.904	35
	2	≥ 25 years	0.710		205
Web	1	No	-0.091	-2.213	119
	2	Yes	0.058		129
Type	1	Public	-0.503	-4.547	59
	2	Private	0.395		177
LTC	1	Yes	-0.370	-1.231	112
	2	No	0.459		141

Table 10. Scoring Calculation

The above table shows that the score increases from the highest risk to the lowest risk category. Thus, it is depicted that the reference category as the first category has the lowest score, while the last category has the highest score. Subsequently, it could be concluded that the reference category matches the assumption that the reference variable is considered as the highest risk than the others.

4.3 Cross-Validation Test

Dataset is validated by utilizing a cross-validation test. Initially, the dataset is separated into training and validation samples. This research uses 100% data for training and 50% data for validation due to the limited data referred to Siddiqi (2006). 50% of data for validation is collected from the top 8 vendor groups – as discussed with Angelise Agri Finance Management Team. Finally, the cross-validation test is conducted by using a default cut-off value of 0.5. The result is shown as follows:

Observed		Predicted		Corrected Rate
		Timely Delivery	Not Timely Delivery	
Training Sample	Timely Delivery	28	23	54.9%
	Not Timely Delivery	4	45	91.8%
Hit Rate				73.0%
Misclassification Rate				27.0%
Validation Sample	Timely Delivery	24	0	100.0%
	Not Timely Delivery	9	21	70.0%
Hit Rate				83.3%
Misclassification Rate				17.7%

Table 11. Cross-Validation Test Using Cut-Off Value of 0.5

The accuracy rate is related to the accuracy of the model's prediction of the timely delivery vendor, while the sensitivity rate is related to the accuracy of the model's prediction of the not timely delivery vendor. According to Table 11, it is shown that the training sample model's accuracy rate of discriminating the vendors with timely delivery is 54.9%, the training sensitivity rate is 91.8%, the overall hit rate in the training sample is 73.0%, and the misclassification rate of 27.0%. Then, in the validation sample, the accuracy rate is inclining to 100.0%, sensitivity rate is declining to 70.0%, hit rate becomes 83.3%, and misclassification rate becomes 17.7%. The high value in hit rate indicates that the model's prediction of the output (assessment of vendors' timeliness of delivery) is 83.3%, and a low value in misclassification rate indicates that the model generates the error of 17.7% in predicting the output.

When the default cut-off value of 0.5 is used, the discriminant capability for vendors with timely delivery is not equal with the vendors with not timely delivery, leading to un-equilibrium discrimination between the two types of vendors (Tsai et al., 2009). Therefore, the optimum cut-off value should be assessed by plotting the hit rate, sensitivity rate, and accuracy rate, with the cut-off value ranges from 0.05-0.95.

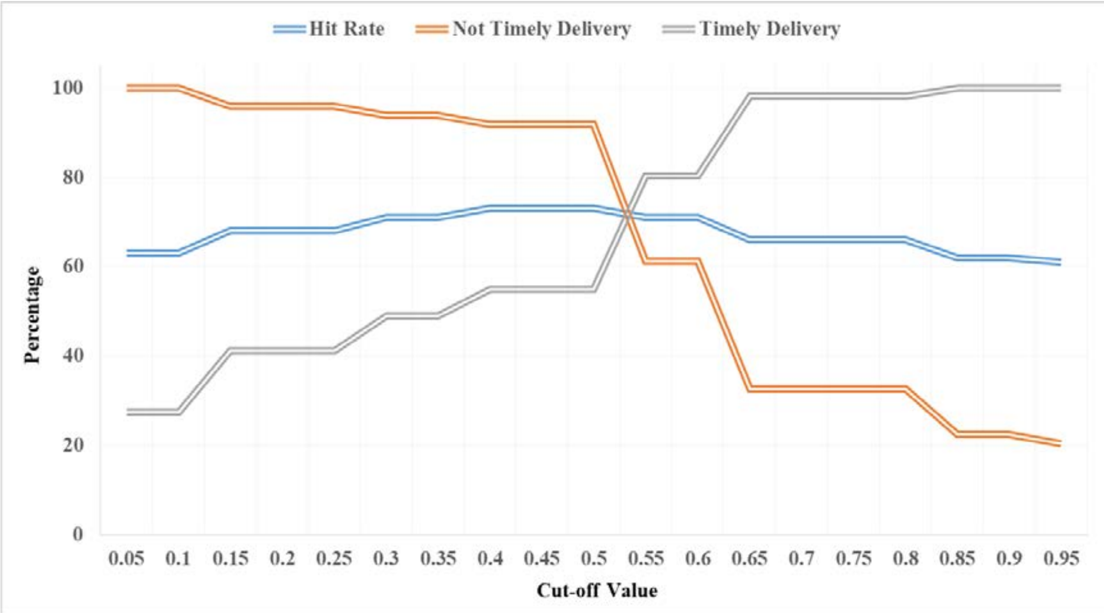

Figure 6. Fitted Cut-Off Value 0.53

According to the cut-off value plot, it is found out that the optimum cut-off value is 0.53, that is, the curve of hit rate, percentage correct of vendors with timely delivery, and percentage correct of vendors with not timely delivery intersect (see Figure 6). The following is the classification table for the cut-off value of 0.53 (Table 12):

Observed		Predicted		Corrected Rate
		Timely Delivery	Not Timely Delivery	
Training Sample	Timely Delivery	28	23	54.9%
	Not Timely Delivery	4	45	91.8%
Hit Rate				73.0%
Misclassification Rate				27.0%
Validation Sample	Timely Delivery	24	0	100.0%
	Not Timely Delivery	9	21	70.0%
Hit Rate				83.3%
Misclassification Rate				17.7%

Table 12. Cross-Validation Test Using Cut-Off Value of 0.53

According to Table 12, it can be seen that the new fitted cut-off value (0.53) generates precisely the same result with the fitted cut-off value of 0.5. Thus, in this research, the higher sensitivity rate serves as a better model, as it acts as the predictor of not timely delivery vendors. Furthermore, a higher sensitivity rate is more helpful in reducing the potency of actual loss than reducing the potency of opportunity loss. Therefore, the final model is the logistic regression with a cut-off value of 0.53.

4.4 Model Comparison

In order to see the improvement and comparison among predicting model that uses complete 13 variables with the model that only uses four variables which fulfill the goodness of fit consideration, the misclassification rate and Area Under Curve (AUC) from each model are compared. Two models that would be compared are as follows:

1. Model 1: Use 13 variables
2. Model 2: Use 4 selected variables that fulfill the goodness of fit consideration

First, the model is compared by using the Receiver Operating Characteristic curve (ROC) to plot sensitivity (true positive) on the Y-axis and (1-specificity) on the X-axis based on the calculation of sensitivity and specificity for all possible cut-off points from 0 to 1. The result of the ROC curve is shown below.

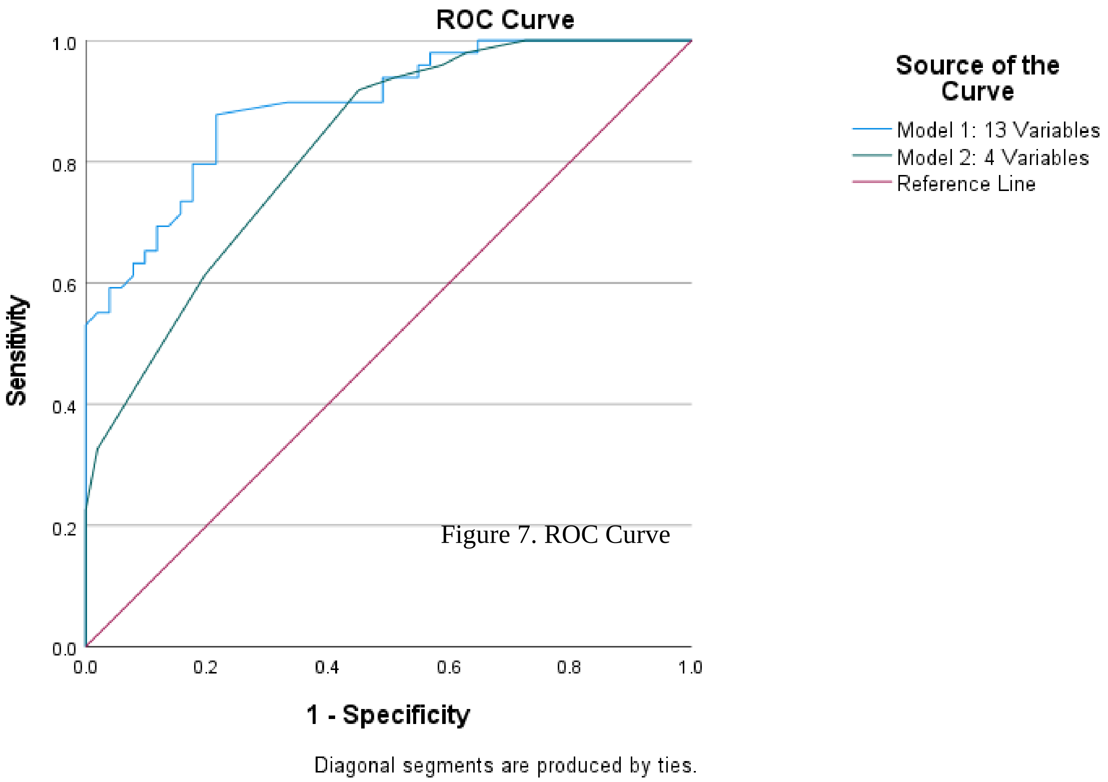

Figure 7. ROC Curve

As depicted in Figure 7 above, there are three lines in the ROC Curve: the purple line indicates the reference line, the light blue line indicates Model 1, and the dark blue line indicates Model 2. According to the figure, it is shown that Model 1 outperforms Model 2. The detailed performance under the ROC Curve is called AUC, indicating the successful rate classification of each model. In addition, the model is compared based on the number of variables used and the misclassification rate of the model cut-off value. The comparison of the models is as follows:

Model	Amount of Variables Used	Area Under Curve (AUC)	Cut-Off Value	Misclassification Rate
Model 1: All Variables	13 Variables	0.892	Model 1 - 0.5	9.3
			Model 1 - 0.485	11.1
Model 2: Selected Variables	4 Variables	0.822	Model 2 - 0.5	17.7
			Model 2 - 0.53	17.7

Table 13. Model Comparison

According to Table 13, Model 1, with a cut-off value of 0.5, has the lowest misclassification rate compared with other models. However, model 1 also has a higher AUC than Model 2, indicating a higher successful classification rate by the model. Using the model with 13 Variables shows that the AUC increases by 7.0% compared to Model 2 with 4 Variables. Although Model 2 has a lower AUC and higher misclassification rate, it is more efficient to choose Model 2 since it only involves four variables within the model. Therefore, it can be concluded that the proposed model to be used is Model 2, using a cut-off value of 0.53.

5. CONCLUSION AND RECOMMENDATION

This research aims to determine the predictability timeliness of delivery of vendors within an agribusiness company. Based on the research, conclusions and recommendations are put forward.

5.1 Conclusion

There are many ways to develop a vendor rating model. In this research, the researchers constructed the vendor rating model by modifying the credit risk scorecard to predict the vendors' timeliness of delivery assessment.

By using the data of 100 vendors from January 2020 to December 2020 and by analyzing credit scorecard and logistic regression, the following conclusions can be drawn:

1. The model fully met the criteria in terms of WOE, IV, and p-value. Thus, the results cannot deviate a lot from the model's assumption. A cross-validation test was performed to check the possibility of model error, and the result is as follows: the optimum has occurred in the fitted cut-off value of 0.53 with the accuracy rate of 100%, sensitivity rate of 70%, hit rate of 83.3%, and misclassification rate of 17.7%. This is the first research to develop a vendor rating model using credit risk scorecard and logistic regression to the researchers' best knowledge.
2. According to the misclassification rate, utilizing all 13 variables in the model minimizes the possibility of misclassification. The AUC demonstrates the improvement of the model, which shows that by using 13 variables, the AUC is increased by 7% compared to the model with only four variables; this shows that the successful classification rate in the predicting model can be improved. Although models that only use four variables have lower AUC and higher

misclassification rates, it is more efficient to choose this model since it only involves four variables within the model.

5.2 Recommendation

Future research is recommended to process data from a more extended period of time with a larger population; in this research, the data is confined to one year only and the top 100 vendors by purchase quantity. The model would be more accurate by taking data from a more extended period and a larger population. To obtain a more specific and complete result, it is also effective to use more independent variables, such as the vendors' production capacity, vendors' ownership structure, vendors' type of business, and vendors' financial health. Having brainstorming with a more significant number of corporate stakeholders might be beneficial in obtaining a diverse perspective of independent variables to be included in the research.

REFERENCES

[1] AberdeenGroup (2002). The Supplier Performance Measurement Benchmarking Report Measuring Supply Chain Success. [online] Available at: http://www.lyonsinfo.com/_resources/aberdeen_spms_report.pdf.

[2] Ampountolas, A., Nyarko Nde, T., Date, P., and Constantinescu, C. (2021). A Machine Learning Approach for Micro-Credit Scoring. Risks, 9(3), p.50.

[3] Azadegan, A. (2011). Benefiting from Supplier Operational Innovativeness: The Influence of Supplier Evaluations and Absorptive Capacity. Journal of Supply Chain Management, 47(2), pp.49–64.

[4] Basel (2000). Principles for the Management of Credit Risk Basel Committee on Banking Supervision Basel. [online] . Available at: https://www.bis.org/publ/bcbs75.pdf.

[5] Franata, R., Faturohman, T. and Rahadi, R.A. (2018). The Implementation of Credit Risk Scorecard Model to Improve the Assessment of Creditworthiness in a Peer-To-Peer Lending Company. International Journal of Accounting, Finance, and Business (IJAFB), 3(13), pp.94–105.

[6] Ghasemi, A. and Zahediasl, S. (2012). Normality Tests for Statistical Analysis: a Guide for Non-Statisticians. International Journal of Endocrinology and Metabolism, [online] 10(2), pp.486–489.

[7] Gordon, S.R. (2008). Supplier Evaluation and Performance Excellence: A Guide to Meaningful Metrics and Successful Results. Ft. Lauderdale, Fl: J. Ross Publishing.

[8] Grant, C. and Osanloo, A. (2014). Understanding, Selecting, and Integrating a Theoretical Framework in Dissertation Research: Creating the Blueprint for Your "House." Administrative Issues Journal Education Practice and Research, 4(2).

[9] Gujarati, D.N. (2004). Econometría. México, D.F.: Mcgraw-Hill.

[10] Khilfah, H.N.L. and Faturohman, T. (2020). Social Media Data to Determine Loan Default Predicting Method in an Islamic Online Peer to Peer Lending. Journal of Islamic Monetary Economics and Finance, 6(2).

[11] Lee, T. and Chen, I. (2005). A two-stage hybrid credit scoring model using artificial neural networks and multivariate adaptive regression splines. Expert Systems with Applications, 28(4), pp.743–752.

[12] Paleologo, G., Elisseeff, A. and Antonini, G. (2010). Subagging for credit scoring models. European Journal of Operational Research, 201(2), pp.490–499.

[13] Siddiqi, N. (2006). Credit Risk Scorecards: Developing and Implementing Intelligent Credit Scoring. Hoboken, N.J.: Wiley.
[14] Tsai, M.-C., Lin, S.-P., Cheng, C.-C. and Lin, Y.-P. (2009). The consumer loan default predicting model – An application of DEA–DA and neural network. Expert Systems with Applications, 36(9), pp.11682–11690.
[15] Veronica, M.S. and Anantadjaya, S.P. (2014). Bankruptcy Prediction Model: An Industrial Study in Indonesian Publicly-listed Firms During 1999-2010. *Review of Integrative Business and Economic Research, 3*(1), 13-41
[16] Wang, Y., Zhang, Y., Lu, Y., and Yu, X. (2020). A Comparative Assessment of Credit Risk Model Based on Machine Learning ——a case study of bank loan data. Procedia Computer Science, 174, pp.141–149.
[17] Wiginton, J.C. (1980). A Note on the Comparison of Logit and Discriminant Models of Consumer Credit Behavior. The Journal of Financial and Quantitative Analysis, 15(3), p.757.
[18] Yap, B.W., Ong, S.H. and Husain, N.H.M. (2011). Using data mining to improve assessment of creditworthiness via credit scoring models. Expert Systems with Applications, 38(10), pp.13274–13283.