SOCIAL MEDIA DATA TO DETERMINE LOAN DEFAULT PREDICTING METHOD IN AN ISLAMIC ONLINE P2P LENDING

Hasna Nabila Laila Khilfah[1] and Taufik Faturohman[2]

[1] SBM Institut Teknologi Bandung, Indonesia, hasna.nabila@sbm-itb.ac.id
[2] SBM Institut Teknologi Bandung, Indonesia, taufik.f@sbm-itb.ac.id

ABSTRACT

Financial technology is growing rapidly in Indonesia. One of main types of financial technologies are online peer-to-peer (P2P) lending platforms. Islamic online P2P lending is also emerging. However, credit risk is still a major concern for this platform. To address this issue, social media assessments have been developed. Therefore, in this paper, the authors have aimed to identify social media variables that could be used as default probability predictors and determine predictability level by adding social media data to the model. Six independent variables consisting of social media data and seven control variables from historical payment and demographic data were used to construct a credit scorecard and logistics. The results identified five variables that could be considered and used as default probability predictors, which are posting frequency at midnight, followers, following, employment, and tenor. Interestingly, the number of religious accounts followed on Instagram is not a significant variable. Furthermore, the model with selected variables through the combination of demographic, historical payment, and social media data could increase the predictability level by 6.6%.

I. INTRODUCTION

1.1. Background

Financial services in Indonesia have changed rapidly due to the development of technology. One common type of financial services, popular with Indonesian citizens, is financial technology. Many arguments have already taken place to define financial technology and to determine exactly what it is. Bank Indonesia (n.d.) defines financial technology as a result of the combination of financial services and technology that changes business models from conventional to moderate ones. Meanwhile, PwC (2016), a global firm focused on audits, assurance, consulting and tax services, determined financial technology or 'fintech' as a dynamic segment between the financial services and technology sector in which people, mostly new entrants, innovate traditional products or services in the financial services industry. Another definition comes from Ancri (2016), which explains that fintech is an industry in which a company uses technology as an efficient way to create a financial system and to deliver financial services. Therefore, it can be concluded that fintech is an innovation within the financial industry in which financial services and technology are combined and lead business model changes from traditional to moderate ones to create financial systems and deliver more efficient financial services.

The rapid development of fintech in Indonesia can be identified by its growth rate and number of transactions. Wibowo, from Bank Indonesia (2017) , explains that the number of fintech transactions in 2017 reached approximately USD 18.6 billion. Furthermore, Statista (n.d.), a global business intelligence portal, which provides online statistics and research, stated that the transaction value of digital payments through fintech's largest segment in Indonesia, would be increased with an annual growth rate of 11.4% (2019–2023) and a projected transaction value as high as USD 50.005 million in 2023. Meanwhile, Indonesia's Financial Services Authority (Otoritas Jasa Keuangan (2018) states that the disbursement of credit through fintech in Indonesia reached IDR 7.8 trillion by the end of July 2018. This means that there was a significant growth of 204.7% compared to the end of 2017, at which time it reached IDR 2.56 trillion.

The rapid development of fintech in Indonesia is supported by opportunities for its market size through the variable of digital market size and the number of unbankable people . With reference to Asosiasi Penyelenggara Jasa Internet Indonesia (2018), the internet penetration rate in Indonesia increased as much as 10.12% between 2017 and 2018, meaning that it reached 64.8% of Indonesia's total population. This opportunity is strengthened by the World Bank (2017), which determines that 6% of the 1.7 billion adults in the world that do not have a bank account live in Indonesia. Thus, it is necessary to find a way to reach unbanked citizens; the answer is to use digital media because the penetration growth rate in Indonesia is promising.

In Indonesia, fintech already has many varieties and is dominated by the payment sector at 42.22% (see Figure 1.1), followed by the lending and aggregator sector (Otoritas Jasa Keuangan & Asosiasi Fintech Indonesia, 2017).

**Profl FinTech di Indonesia
(Berdasarkan Sektor)**
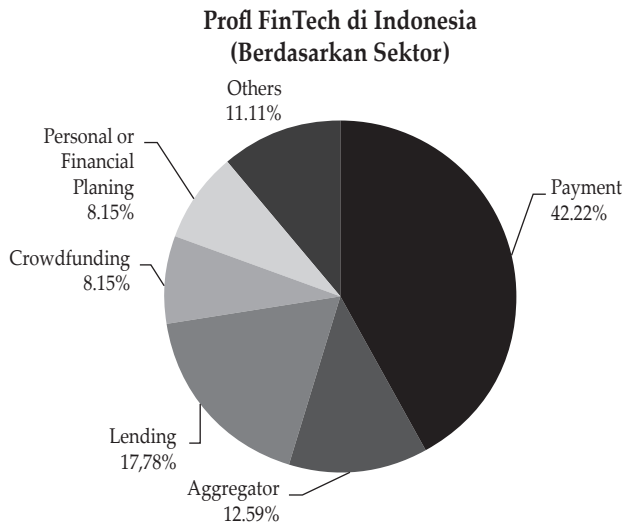
Source: "Fintech di Indonesia" by Asosiasi Fintech Indonesia and Otoritas Jasa Keuangan (2017)

**Figure 1.
Fintech Profile in Indonesia Based on Sectors**

In September 2018, 67 fintech lendings had been registered in Otoritas Jasa Keuangan (OJK) with a combination of 65 conventional and two Islamic peer-to-peer (P2P) lending types (Otoritas Jasa Keuangan, 2018), followed by 40 fintech lending companies, which are still undergoing the registration process.

Within their development, financial institutions have faced many challenges and risks, especially credit risk. According to the PwC Indonesia Banking Survey (2018) , credit risk is the top challenge for banks, as major financial institutions, to grow. It is stated that even though the non-performing loan (NPL), the bad debt credit ratio, was expected to decline in 2018, credit risk is still a major concern. OJK (2018) even noted that credit risk is still maintained as it is reflected in the NPL of the bank to as much as 2.61%, while the non-performing financing (NPF) of Islamic financial institutions is 4,41%, which is below the maximum rate of 5%. Thus, it is clear that credit risk is still a major concern in Indonesia even though it is expected that it will still be maintained as the NPL and NPF are below 5%.

To manage credit risk, some studies started using big data analysis, utilising social media as default predictors. Tan and Phan (2016) showed that social networks derived from social media data could increase the default predictability level within a microfinance institution. Chen et al. (2016) also discovered that social media data, including demographics, tweets and network data could have a better discriminating power between good and bad debt borrowers in P2P lending companies. However, in Indonesia, several regulations need to be considered for the usage of data in social media, such as 'Undang-Undang No. 11 Tahun 2008 " article 26 about "Informasi dan Transaksi Elektronik" and "Peraturan Pemerintah No. 82 Tahun 2012" about "Penyelenggaraan Sistem dan Transaksi Elektronik" stated that data usage must have consent for the related individual.

This research will use the case from an online P2P lender in Indonesia. Built in March 2017, it faced a fluctuation in its non-performing financing between May and September 2018. It is already known that the lending platforms is the second largest fintech in Indonesia; it is still exposed to credit risk even though the process has been made efficient through the use of technology and the internet. To address this issue, the company is developing a social media assessment. Thus, based on previous studies related to the company's issue, this research will validate whether or not the social media assessment done by the company is effective to increase its credit scoring predictability level with selected variables from previous studies.

It is expected that the research findings will be considered by the company to build a strategy in relation to its credit scores; this will help the company improve its performance by minimising its credit risk through its improvement in its credit scoring, especially in relation to social media assessment. Even though the company is still in its early establishment period, it could compete in the fintech industry, especially in online P2P lending services.

1.2. Objective

Based on the problem statement above, the research objectives of this research are:
1. To identify the social media variables that could be predictors for the Islamic online P2P lending company's credit scores
2. To determine the predictability level of credit scoring using the social media data of an Islamic online P2P lending company

II. LITERATURE REVIEW
2.1. Background Theory
2.1.1. Online P2P Lending Development

P2P lending could be defined as a new online platform that provides an alternative method of applying for credit (World Bank Group, 2018a) . This platform has a faster process for loan applications due to its replacement of traditional scoring credit models with machine learning. It also, typically, provides shorter-term loans than those awarded by banks. A definition of online P2P lending is also given in Peraturan Otoritas Jasa Keuangan Nomor 77/POJK.01/2016, which explains it as a financial service used to connect borrowers and lenders and to generate electronic rupiah lending agreements via the internet. P2P lending can be defined as a platform that utilises digital media to provide capital by connecting lenders and borrowers. P2P lending also increases the speed of the process.

Online P2P lending has grown rapidly around the world. In 2016, the amount of credit extended using this platform increased from 11 billion USD in 2013 to 284 billion USD (Claessens, Frost, Turner & Zhu, 2018). Furthermore, a significant increment of the P2P lending transaction value had emerged between 2015 and 2016, which meant that China, the United States, and the United Kingdom have the largest lending platform markets (World Bank Group, 2018b).

The development of online P2P lending has also emerged in Indonesia, where there are opportunities for platform development. These opportunities are supported by the growing number of unbankable small and medium enterprises

(SMEs), which account for 60%–70% of the 56.4 million SMEs in 2014 (Bank Indonesia, 2015). Furthermore, Indonesia accounts for 6% of the global figure of 1.7 billion unbanked adults (those who do not have a bank account) (World Bank, 2017). The number of unbankable citizens and SMEs are a potential, yet promising, market that could be penetrated by online P2P lending through digital or network platforms. The growth of online P2P lending is clear from the total lending accumulation between January and September 2018; during this time, lending increased from 3 trillion rupiah to 13 trillion rupiah (Otoritas Jasa Keuangan, 2018). In September 2018, 67 fintech lending platforms were registered in OJK per with a combination of 65 conventional and two Islamic P2P lending platforms (OJK, September 2018). This will be followed by 40 fintech lending companies, which are still undergoing the registration process. From viewing the opportunities and significant growth, it can be expected that online P2P lending will undergo promising development in Indonesia.

2.1.2. Islamic Online P2P Lending

Islamic online P2P lending is emerging in Indonesia. This platform is offered to improve Islamic banking systems by increasing competitiveness, minimising overhead costs, and minimising credit gaps by reaching the younger generation and refining its strong image (Todorof, 2018) . Furthermore, Islamic online P2P lending could benefit consumers by providing return without additional costs; this could make it more convenient for Muslim customers to apply Islamic principles on lending platform (Novalia, 2018). Therefore, it can be concluded that the emergence of Islamic online P2P lending is driven to minimise overhead costs and improve Islamic financial institution competitiveness by offering additional benefits to customers, in particular Muslim customers.

A study related to religion-related credit scoring could be applied to Islamic online P2P lending; the objective of the research is to explore Islamic online P2P lending, which had already applied an assessment in relation to the religious aspect. Tan and Phan (2016) used religion affiliation as user preference in social media to determine the likelihood of defaulting. This paper will use religious preferences as one of its independent variables, but considering the number of religious accounts followed on social media. Kosinski et al. (2013) state that religious preferences (and 'likes') on Facebook could be a predictor of human personal traits and attributes.

This paper uses Islamic online P2P lending as its research objective with the following characteristics

a. The company uses Murabahah as an agreement principal for which the sellers purchase things that are requested by the buyers and are priced with an agreed margin.
b. The company gains profit from the margin, so the price would be more expensive.
c. For transactions using Islamic online P2P lending, two stages are required; first, Wa'ad (ordering things from a description given by a buyer) and second, Murabahah, where there is an agreement about the selling price between sellers and buyers.

d. The company uses its website and applications as transaction media and develops a technological integrated system to assess and process the buyers and their transactions.

2.1.3. Credit Risk

Credit risk is the potential that a bank borrower will fail to meet his obligations in relation to the agreed terms (Basel, 2000). Meanwhile, credit risk management is a structured approach to managing uncertainties through risk assessment; it aims to maximise the bank's risk-adjusted rate of return by keeping the credit risk within favourable parameters (Greuning & Iqbal, 2007; Nyong'O, 2009. The analysis from credit risk management results in positive credit risk decisions (Brown & Moles, 2014). Financial institutions can exposed by credit risk; this leads them to undergo credit risk management by analysing credit when consumers wants to assign credit to the firm and maintaining the credit risk within acceptable parameters to minimise risk and manage uncertainties.

This paper will focus on the development of a credit scoring model because this is something that P2P lending companies already have. Furthermore, credit scoring is also included as one of the main techniques for assessing credit risk within financial institutions.

2.1.4. The Credit Scoring Model

Credit scoring is an essential way of recognising different groups in a population when one can only see related characteristics and not those that separate the groups. It offers a rating system that is formalised into a mathematical or statistical model and is based on multivariate statistical inference techniques, which use the same data and methodology to assess the credit (Brown & Peter, 2014; Thomas, 2000). The benefits of using this model are more precise and transparent than other rating systems that still depend on judgement. Moreover, credit scoring analyses a large sample of good and bad credits to create a differential equation. The popular method in this model is the Z-score method, which predicts the probability of failure based on discriminant analysis techniques.

Previous studies about credit scoring have been conducted. It is proven that certified information provided by the platform is the most important determinant for the credit scoring model and private information in the form of credit scores can predict loan performance better than public information. In this research, private information refers to information about property, income and historical credit records. Furthermore, loans with a lower credit grade and longer duration are more likely to have a higher default rate (Chen, Li, & Zeng, 2018; Emekter et al., 2014), which indicates that loans with lower credit grades and longer payment terms are likely to have a higher default rate. Lending Club is a popular P2P lending platform in the United States; as a research object with data about requested loan amounts, it uses the number of recent credit inquiries, credit history length, total open credit accounts, currently open credit accounts, and revolving line utilisation to determine the final grade used to determine the interest rate on the loan. Therefore, the credit scoring model is essential for predicting the

probability of a borrower's loan default; this means that, if using the credit scoring model, financial institutions could take an optimal decision to avoid losses that could come from such credit risk exposure.

2.1.5. Behavioural Scoring

Behavioural scoring could support decisions regarding the management of existing consumers by forecasting their performances (Thomas, 2000). Decisions could be made in relation to how much credit should be assigned. This method develops a longitudinal forecasting system using cross-sectional data with experimentation, usually over a 12-18-month period. However, periods as short as 6 months can be sufficient to build a scoring system to predict behaviours that could cause consumers to default. Behavioural scoring could also quantify consumer behaviour; through this model, firms would be able to better understand their customers and facilitate customer management decisions (Scorto, n.d.). This study will focus on behavioural scoring when data is mainly derived from existing customers.

2.1.6. Social Credit Score

Dellarocas et al. (2016) interpret social credit score as exploring customer creditworthiness through social network information and taking a broader view of consumers' social status. Tan and Phan (2016) prove that added social network data could increase predictability rates by 18%. Social networks and status could be defined through social ties and P2P interaction on the borrower's social media site. Social credit scoring is expected to help with the scoring of individual loans. Therefore, using data derived from borrower's social network data is expected to make the credit scoring model more accurate and precise. This study will focus on the social aspect of borrowers through their social media accounts to determine their likelihood of defaulting on their credit.

2.1.7. Loan Default Predicting Method

Tsai et al. (2009a) define the loan default predicting method as a way to analyse historical information about credit consumers to predict whether they will be able to pay back their loans. To predict the defaulting of a borrower, several methods can be used; these are already applied in some research. For instance, Tsai et al. (2009b) analysed customers of an unsecured consumer loan from a financial institution in Taiwan by using and comparing four methods: Discriminant Analysis, DEA-Discriminant Analysis (DEA-DA), Logistic Regression (LR), and Neural Network (NN). They determined that DEA-DA and NN analysis have better-predicting capability and are an optimal predicting model for this research. Hsieh and Lee (2011) analysed borrowers from a Taiwan bank using and comparing three methods: Linear Discriminant Analysis (LDA), Backpropagation Neural Networks (BPN), and the Support Vector Machine (SVM). The result of this research indicates that BPN has better accuracy and is an efficient method for behavioural scoring. Another study comes from Tan and Phan (2016), who analysed the social-network-based credit scoring method for microloans. They

used the logistic regression and random forest methods to compare the results between the addition of network metrics and customer preferences as predicting variables, although the model does not use both of the variables. The results show that social network data could increase predictability measurements by 18%.

2.2. Previous Studies

Methods exist that could be used to determine predictability levels in credit scoring: logistic regression (as the common method), decision trees, credit scorecards, neural networks and random forest. However, this paper will use credit scorecards to develop the research model. The development of the credit scorecard model is explored in a study by Yap et al. (2011); their study compares three methods – scorecard, logistic regression, and decision tree – to determine predictability levels with demographic data as the dependent variables. The results determined that the scorecard method has the lowest misclassification rates compared to the other two methods. Furthermore, the scorecard model is more applicative because it could eliminate unnecessary variables. Therefore, the development of the scorecard model could be suitable for efficient use in a company; this would enable the company to focus on the most important variables for the development of the social media assessment.

The variables used in this paper were considered by previous studies to be more suitable and appropriate for the company. The social media data, as the independent variable, were derived from Chen et al. (2016), which comprised the duration of social media usage, the number of borrowers posting within a certain amount of time, the number of posts, and the number of followers or people following. The social media platforms used were Instagram and Facebook, while Chen et al. (2016) used Weibo, because this company is already focused on these two social media platforms. Because the study is related to Islamic P2P lending, it also used the number of religion accounts affiliated to borrowers on social media, in accordance with the previous study by Tan and Phan (2016). Demographic variables and control variables were also used; these were derived from Tsai et al. (2009) and Yap et al. (2011).

III. METHODOLOGY

This research aimed to evaluate the impact of demographics, historical payment, and social media data in improving the predictability level of an online Indonesian P2P lender's credit scoring model; this was done to help P2P lending companies avoid credit losses and utilised the scorecard and logistic regression models. This chapter will determine and identify the steps taken to answer the research question and reveal relevant factors.

3.1. Data

This paper obtained its data from secondary data sources such as journals, websites, books, and from the company's borrowing data from March 2017–January 2019. Primary data was taken from social media platforms with a focus on Instagram and Facebook due to the company already utilising these platforms. This study

could only use available data as some data was limited; not all of the borrowers have social media accounts and some accounts were locked for privacy. The social media data was filtered and t 21 posts was the minimum number of posts that were not linked to privately locked accounts (Chen et al., 2016). Due to the opinions of Parveen and Showkaat (2017), convenience sampling was used; they stated that convenience sampling is appropriate for use under conditions in which the object is accessible with criteria that must be met. From a total of 465 borrowers, only the accounts of 107 could be accessed; this means that this research will only use those 107 borrowers. This data is being used because it is already proven in other previous studies and consists of one binary dependent variable combined with categorical and numeric dependent variables, which are suitable for the use of logistic regression to develop the credit scoring model.

3.1.1. Operationalisation of Variables

This paper has three variables: independent, control, and dependent. The independent variables have an impact on the dependent variable, which will be tested. According to the previous research by Yap et al. (2011), Tan and Phan (2016), Hsieh et al. (2011), and Tsai et al. (2009), this study will use social media data as an independent variable, while the probability of defaulting acts as dependent variable. Furthermore, control variables are comprised of demographic and historical payment data. The demographic data is comprised of age, gender, marital status, education, employment, income, and district (Tsai et al., 2009; Yap et al., 2011). These seven control variables are involved because they are already proven to be significant in terms of the probability of default as dependent variables; they are also being used by companies for their credit assessments. Meanwhile, historical payment data consists of the instalment amount and tenor. Furthermore, this study would like to obtain data that comprises the duration of usage on Instagram (IG_Month), how much borrowers post on Facebook, i.e. midnight, morning, afternoon, evening, or night (Posting_Midnight, Posting_Morning, Posting_Afternoon, Posting_Evening, Posting_Night), how many followers they have on Instagram (Followers), how many people they are following on Instagram (Following), how many of the total number of posts divided by the number of months duration on Instagram (IGPost_Month) (Chen, et al. 2018), and number of followed or affiliated religious accounts (Tan & Phan, 2011). Therefore, the total number of independent variables is ten, with nine control variables. If the variable has been assessed through the weight of evidence (WoE), information value, and p-value from the initial logistic regression, then the equation could be constructed as follows:

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

Where, x_i is the variable that is selected for inclusion in the credit scoring model. The variables used in the study are shown in Table 1.

Table 1.
Operationalisation of Variables

Variable Name	Description of Variable	Role
Creditworthiness	The output of customer's creditworthiness assessment (default/non-default)	Dependent
Gender	Borrower's gender	
Age	Number of age in years	
Marital status	Marital status	
Education	Last education stage of borrowers	
Employment	Type of work	
Income	Amount of money that borrowers get from their monthly employment	
District	City where borrowers live	Control
Tenor	Credit duration	
Instalment	The amount of money that has to be paid for the loan	
Followers	Number of followers on Instagram	
Following	Number of people following on Instagram	Independent
Account Usage Duration	The user's duration using Instagram according to the date of their first posting in the month	
Number of Post per Month	The proportion between total number of posts on Instagram and how long the Instagram account is used per month	
Religion Preference	The number of religious accounts that are followed by the borrower on Instagram	
Posting at Midnight	The number of Facebook posts at midnight (00.00-03.59) from the last 21 posts of each account	Independent
Posting in the Morning	The number of Facebook posts at midnight (04.00-10.00) from the last 21 posts of each account	
Posting in the Afternoon	The number of Facebook posts in midnight (10.01-14.00) from the last 21 posts of each account	
Posting in the Evening	The number of Facebook posts at midnight (14.01-18.30) from the last 21 posts of each account	
Posting During the Night	The number of Facebook Posts at midnight (18.31-23.59) from the last 21 posts of each account	Independent

Source: Researcher Data Analysis

3.2. Model Development

The model will be advanced using scorecard development; this was derived from Yap et al. (2011) because it can directly focus on the most important variables through its consideration of predictive power and risk for each dependent variable. During scorecard development, logistic regression is involved as follows:

$$log\left(\frac{p}{1-p}\right) = w_0 + \sum w_i log x_i \tag{2}$$

Where, x_i is the dependent and control variables that are selected to be included in the credit scoring model.

The variables used were considered in the previous studies and were adjusted to the current social media assessment already developed by the company. In line with previous studies, this paper constructs a hypothesis for each independent variable. There are ten independent variables comprising social media data relating to the duration of Instagram usage in the form of months (IG_Month), how many borrowers posted on Facebook at midnight, in the morning, during the afternoon, in the evening, and during the night (Posting_Midnight, Posting_Morning, Posting_Afternoon, Posting_Evening, Posting_Night), how many followers on Instagram (Followers), how many following on Instagram (Following), how many posts divided by the number of months on Instagram (IGPost_Month) (Chen et al., 2016) , and how many religious accounts are followed on Instagram (Rel_account) (Tan & Phan, 2016). Meanwhile, nine variables were ruled as control variables: age, gender, marital status, employment, education, district, income (Tsai et al., 2009; Yap et al., 2011), tenor, and number of instalments (Hsieh et al., 2011). Those independent and control variables were combined from continuous and dummy or categorical data. The control variables were needed to make the logistic regression method more accurate. Moreover, the dependent variable in this research is one that relates to the borrower's tendency to default. These variables were used because they are commonly found in previous research to predict the likelihood of borrowers defaulting on their loan payment, especially within a fintech company with P2P lending; it is suitable for the current social media investigation conducted by the company.

Regarding previous studies, this paper aims to determine whether social media data could increase the predictability level of credit scoring or not. Therefore, the hypothesis could be constructed as follows:

H_0 = Social media data cannot increase the predictability level of credit scoring
H_1 = Social media data can increase the predictability level of credit scoring

3.3. Method

To help process the amount of data that exceeds 100, the data was processed using SPSS, STATA, and Microsoft Excel; this made the processing time more efficient. Before the scorecard development was conducted, the data was tested for its multicollinearity and normality. The credit scorecard was used as the research method because it could focus directly on the appropriate and important variables to make the assessment more efficient for the company.

The scorecard development began by calculating the risk and predictive power of each variable according to their WoE and information value to eliminate variables that did not fulfil the requirements. Next, logistic regression was developed to determine the conditional probability of a specific applicant being a defaulter or non-defaulter; here, the generated coefficient would also be used for the scoring calculation. Furthermore, scoring calculations began to see the trend from the lowest to the highest risk attributes for each variable. The developed model was then tested to assess its goodness of fit. The details of the method analysis are discussed in the next section.

3.3.1. Multicollinearity Test

Multicollinearity is a statistical phenomenon in which two or more predictor variables in a multiple logistic regression model are highly correlated or associated (Midi et al., 2013a). It is stated that multicollinearity does not reduce the predictive power of the model as a whole, but it could make the optimal variable selection very difficult in standard logistic regression. Furthermore, Siddiqi (2005) states that multicollinearity is not a major factor when a model comprises a large number of datasets. However, this research only has 107 datasets. Thus, it was important that the multicollinearity test was conducted to ensure that no predictor variables correlated with each other. Midi et al. (2013b) explains that multicollinearity can be utilised to test the tolerance (TOL) and variance inflation factor (VIF). The rule of thumb is that if the VIF of a variable exceeds 10, this indicates a high collinear level; if the tolerance level is closer to zero, it indicates a greater degree of collinearity of that variable with other variables.

3.3.2. Variable Selection & Reduction
a. Weight of Evidence (WoE)

Yap et al. (2011) defines the WoE of an attribute as a logarithm of the proportion of 'goods' in the attribute over the proportion of 'bads'. The values will have resulted in a high negative or a high positive. High negative values correspond to high risk, while high positive values correspond to low risk. Bolton (2009a) states that WoE is used to assess the relative risk of different attributes and to transform the character into the variable. Moreover, WoE variables have a linear relationship with the logistic function, so it makes the WoE more appropriate when it represents the characteristic using logistic regression. The formula of the WoE of attributes is as follows (Bolton, 2009b):

$$Weight\ of\ Evidence\ _{attribute} = log\ \frac{p_{goodattribute}}{p_{badattribute}}\ where \quad (3)$$

$$p_{badattribute} = \frac{\#goods_{attribute}}{\#goods}\ and$$

$$p_{badattribute} = \frac{\#bads_{attribute}}{\#bads}$$

b. Information Value

Information value is required to support the variable selection that is included in a credit scoring model; this is done by assessing its variable predictive power and its ability to separate high risks from low risks . Information value is determined through the weighted sum of the WoE of the variable's attributes. The weight is the difference between the good proportion and bad proportion with respect to the attributes. The information value should be greater than 0.02 if a variable is to be considered in the credit scoring model (Yap et al., 2011). Furthermore, information values lower than 0.1 can be considered as a weak predictive power; scores smaller

than 0.3 are associated with medium predictive power, those smaller than 0.5 are strong, and those greater than 0.5 may mean that the characteristics are being over predicted. The information value is computed as follows (Bolton, 2009):

$$Information\ Value = \sum \left(p_{goodattribute} - p_{badattribute} \right) * Weight of Evidence$$

c. Logistic Regression

Logistic regression is a popular technique used for credit scoring, in which the probability of a dichotomous outcome (Y=0 or Y=1) is related to set potential predictor variables. In credit scoring, the objective of this method is to determine the conditional probability of a specific applicant being a defaulter or non-defaulter, in relation to the values of the independent variables of that credit applicant (Yap et al., 2011). Logistic regression was used in this study because the dependent variable is categorical (uses non-metric data), but the independent variable has a mixture of continual (metric data) and categorical variables (non-metric data). Furthermore, logistic regression is a standard statistical technique that aims to estimate the probability of loan default and borrowers' characteristics based on historical data (Tabagari, 2015).

3.3.3. Scoring Calculation

With reference to Yap et al. (2011) and Siddiqi (2005), to calculate the score, scorecard points must initially be calculated as follows:

Scores = log (odds) * factor + offset
Factor = points to double odds/log 2,
Offset = score – factor * log (odds)

The score can then be calculated as follows (Siddiqi, 2005):

$$-\left(woe_i * \beta_i + \frac{\alpha}{n} * factor + \frac{offset}{n} \right) \tag{4}$$

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

3.3.4. Predicting Model Goodness of Fit
a. Cross Validation Test

Tsai (2009) cited from West (2000) determines that cross-validation is frequently applied to construct the credit scoring model, so it can reduce the mutual influencing effect among the samples and upgrade the reliability of the assessment

results. The dataset was divided into training and validity samples. A training sample was used to estimate the credit scoring model's parameter, while a validity sample was used to test the capability of the built model (Hsieh et al., 2011). Following previous research by Yap et al. (2011) the dataset was divided as follows: 70% for the training sample and 30% for the validity sample. However, Tsai et al. (2009) divided the dataset into 66% for the training model and 33% for the validity sample, with random sampling; thus, to construct the model, this research is divided into two sets, the higher portion reflecting the training sample.

A cross-validation test involves the analysis of the classification table. Tabagari (2015) defines this as a table that contains observed and predicted model results or results from training and validity samples. The data record was classified using the computed probability given by the model and cut value that is the minimum probability value that should be classified as 1. Generally, the default cut-value is 0.5, which means that each dataset that has a value larger than 0.5 should be classified as 1. We then compared the predicted target variable with an observed value, so we could measure and evaluate the predictive accuracy of the non-default borrower accuracy rate; it also accurately predicted the default borrower's sensitive rate and the model predicting accuracy's hit rate.

b. Receiver Operating Characteristic Curve (ROC Curve)
Tabagari (2015) explains the receiver operating characteristic (ROC) curve as one that plots 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 area under the ROC curve ranges from 0.5 and 1.0 with larger values indicating a better fit. The ROC curve is also called a trade-off curve because it shows the trade-off between 'goods' and 'bads'; it refers to the percentage of total 'bads' that must be accepted to accept a given percentage of total 'goods'. Meanwhile, Bolton (2009) states that the area under the ROC curve (AUC), so called as the model's predictive power in credit scoring, measures the model's ability to discriminate between those observations that experience the outcome interest and those observations that do not. The AUC is a value that varies from 0.5 (a discriminating power that is no better than chance) to 1.0 (a perfect discriminating power).

IV. RESULTS AND ANALYSIS
4.1. Results
4.1.1. Assumption Test (Normality and Multicollinearity)
Initially, this paper used age and education for the independent variables; however, there are missing values so only 82 datasets remain. So, the normality test was conducted using 82 datasets with the following results (see Table 2):

Table 2.
Normality Test

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	**Df**	**Sig.**	**Statistic**	**Df**	**Sig.**
Late	.446	82	.000	.571	82	.000
Age	.126	82	.003	.935	82	.000
Gender	.470	82	.000	.533	82	.000
Marital status	.428	82	.000	.593	82	.000
District	.274	82	.000	.853	82	.000
Employment	.302	82	.000	.837	82	.000
Education	.316	82	.000	.828	82	.000
Income	.216	82	.000	.664	82	.000
Instalment	.221	82	.000	.677	82	.000
Tenor	.328	82	.000	.822	82	.000
IG_Months	.108	82	.019	.982	82	.301
Posting_Midnight	.307	82	.000	.682	82	.000
Rel_Account	.279	82	.000	.627	82	.000
Followers	.218	82	.000	.608	82	.000
Following	.225	82	.000	.573	82	.000
IGPost_Month	.195	82	.000	.770	82	.000

Source: Researchers' Data Analysis

The normality test shows that no variable has a significance level above 0.05, which means that the data is not normal; therefore, we decided not to use the Age and Education variables to increase the number of datasets to 107. Furthermore, outliers were checked as a data preparation step and the decision was made to remove the outliers from the dataset. Therefore, the final number of datasets that were used in the research was 100, which was considered as normal (Gujarati, 2004).

After eliminating the Age and Education variables, the next step was to conduct a multicollinearity test to ensure that each of the independent variables were not correlated. The multicollinearity test utilised the variance inflation factors (VIF) test and achieved the following results (see Table 3):

**Table 3
Initial Multicollinearity Test**

Model	Standardised Coefficients	T	Sig.	Collinearity Statistics	
	Beta			Tolerance	VIF
(Constant)		-0.500	0.619		
Gender	-0.025	-0.273	0.785	0.873	1.145
Marital status	-0.150	-1.514	0.134	0.750	1.333
District	0.091	0.889	0.376	0.708	1.412
Employment	0.107	1.157	0.250	0.859	1.164
Income	-0.090	-0.765	0.446	0.534	1.874
Instalment	-0.099	-0.826	0.411	0.515	1.940
Tenor	0.420	3.986	0.000	0.666	1.501
IG_Month	-0.116	-1.190	0.237	0.776	1.289
Posting_Morning	0.749	0.504	0.616	0.003	298.477
Posting_Afternoon	0.641	0.553	0.582	0.006	181.509
Posting_Evening	0.551	0.480	0.632	0.006	177.925
Posting_Night	0.729	0.516	0.607	0.004	269.335
Posting_Midnight	0.586	0.852	0.396	0.016	63.787
Rel_Account	-0.028	-0.299	0.765	0.859	1.165
Followers	-0.117	-1.148	0.254	0.714	1.401
Following	0.195	2.067	0.042	0.829	1.207
IGPost_Month	0.133	1.424	0.158	0.852	1.174

Source: Researchers' Data Analysis

From the VIF test, it was clear that multicollinearity exists – tolerance score < 0.1 and variance inflation factors > 10 – within the variable of posting frequency in the morning, afternoon, evening, night, and midnight. The decision was made to only use the data from the posting frequency at midnight because Chen et al. (2016) assume that people who post at midnight are more likely to default on their loan. Another multicollinearity test was then conducted, the results from which are exhibited in Table 4:

Table 4
Revised Multicollinearity Test

Model	Standardised Coefficients	T	Sig.	Collinearity Statistics	
	Beta			Tolerance	VIF
(Constant)		1.483	0.141		
Gender	-0.084	-0.860	0.392	0.902	1.108
Marital status	-0.111	-1.100	0.274	0.839	1.192
District	-0.073	-.720	0.474	0.835	1.197
Employment	-0.118	-1.215	0.228	0.903	1.107
Income	-0.052	-0.422	0.674	0.564	1.774
Instalment	-0.005	-0.040	0.968	0.573	1.744
Tenor	0.316	2.835	0.006	0.689	1.452
IG_Month	-0.127	-1.221	0.225	0.796	1.257
Posting_Midnight	0.198	1.851	0.067	0.747	1.338
Rel_Account	0.045	0.440	0.661	0.830	1.205
Followers	-0.103	-0.944	0.348	0.726	1.378
Following	0.059	0.593	0.555	0.874	1.144
IGPost_Month	0.002	0.018	0.986	0.885	1.130

Source: Researchers' Data Analysis

The results show that all of the variables have a tolerance score higher than 0.1 and the variance inflation factors (VIF) < 10. This indicates that no multicollinearity exists or that the independent variables do not correlate; therefore, the decision was made not to use the independent variables in the research.

4.4. Information Values
To determine the predictive power of each variable, their information value was calculated with the following results (see Table 5):

Table 5
Information Value

Variable	Information Value	Ordering	Predictive Power	Decision
Instalment	0.603	1	Over-Predicting	Input
Tenor	0.395	2	Strong	Input
Employment	0.221	3	Strong	Input
Following	0.195	4	Medium	Input
Income	0.177	5	Medium	Input
IG Post per Month	0.165	6	Medium	Input
Posting in Midnight	0.160	7	Medium	Input
IG Duration in Month	0.130	8	Medium	Input
Religion Account	0.108	9	Medium	Input
Followers	0.107	10	Medium	Input
District	0.062	11	Weak	Reject
Gender	0.056	12	Weak	Reject
Marital status	0.044	13	Weak	Reject

Source: Researchers' Data Analysis

The results in Table 5 show that three of the variables must be eliminated because they have a weak predictive power or information value of less than 0.1. Furthermore, Instalment, which is the only variable with over-predicting power or information value of more than 0.5, was still included because the Instalment WoE showed the logical trend order and over-predicting power variables were still included in the model constructed by Yap et al. (2011). Therefore, 10 were included as the input for information value decision; these are ordered from highest to lowest predictive power, respectively: Instalment, Tenor, Employment, Following, Income, Instagram posts per month, Posting at Midnight, Instagram usage during a month, Religious account, and Followers.

4.5. Logistic Regression
4.5.1. Initial Logistic Regression
Initially, logistic regression was conducted for all variables so that the p-value or significance level for each variable could be considered for goodness of fit. The discrete variables were directly categorised because, according to Yap, et al. (2011) and Siddiqi (2005), in scorecard development, variables should be categorised according to the logical trend from the WoE result. These initial logistic regression results can be seen in the following (See Table 6):

Table 6
Initial Logistic Regression

		Coef.	Std. Err.	Z	P>|z|
Step 1a	Gender	-0.186	0.692	-0.27	0.788
	Marital status	-1.169	0.812	-1.44	0.150
	District	-0.020	0.397	-0.05	0.960
	Employment	-0.843	0.339	-2.49	0.013
	Income	0.079	0.312	0.25	0.799
	Instalment	0.257	0.214	1.20	0.231
	Tenor	-1.734	0.6663	-2.61	0.009
	IG_Month	0.026	0.375	0.07	0.944
	Posting_Midnight	-2.606	1.188	-2.19	0.028
	Rel_Account	-0.493	0.444	-1.11	0.268
	Followers	-0.714	0.337	-2.12	0.034
	Following	-1.109	0.377	-2.95	0.003
	IGPost_Month	-1.170	0.622	-1.88	0.060
	Constant	16.416	4.563	3.60	0.000

Source: Researchers' Data Analysis

According to Table 5, only five variables have a significance level of less than 0.05 (p-value < 0.05): Employment (0.013), Tenor (0.009), Posting in Midnight (0.028), Followers (0.034), and Following (0.003). Furthermore, there is a reduction in the log likelihood or error with a significance level of 0.0001 < 0.05, which means that the declining error is significant.

Logistic regression			
	Number of obs	=	100
	LR chi2 (13)	=	42.06
	Prob > chi2	=	0.0001
Log likelihood = -40.054	Pseudo R2	=	0.3443

Source: Researchers' Data Analysis

Figure 2
Log Likelihood of Initial Logistic Regression

4.5.2. The Goodness of Fit

The next step was to analyse each independent variable according to the goodness of fit criteria, comprised of WoE, information value, and p-value. The analysis determined whether or not the independent variables should be included (see Table 7):

Table 7
The Goodness Fit Criteria

Variable	WOE	Information Value	p-Value	Decision
Gender	Yes	No	No	No
Marital status	Yes	No	No	No
District	No	No	No	No
Employment	Yes	Yes	Yes	Yes
Income	No	Yes	No	No
Instalment	No	Yes	No	No
Tenor	Yes	Yes	Yes	Yes
IG_Month	Yes	Yes	No	No
Posting_Midnight	Yes	Yes	Yes	Yes
Rel_Account	Yes	Yes	No	No
Followers	Yes	Yes	Yes	Yes
Following	Yes	Yes	Yes	Yes
IGPost_Month	Yes	Yes	No	No

Source: Researchers Data Analysis

From the consideration of WoE, from considering WoE, information value, and p-value, Table 7 shows five variables that fulfil those three aspects; they can be included in the scorecard because they represent the borrower's creditworthiness. The five variables are: Employment, Tenor, Posting at Midnight, Followers, and Following.

4.5.3. Logistic Regression Model

The final logistic regression model was constructed from the selected independent variables and the following results were achieved (See Table 8):

Table 8
Logistic Regression Model

	Coef.	Std. Err.	z	P>|z|	Exp (B)
Employment	-0.645	0.288	-2.24	0.025	0.524
Tenor	-1.570	0.543	-2.89	0.004	0.208
Posting_Midnight	-2.139	1.001	-2.14	0.033	0.118
Followers	-0.739	0.285	-2.60	0.009	0.478
Following	-0.984	0.331	-2.97	0.003	0.374
Constant	11.556	2.972	3.89	0.000	

Source: Researchers Data Analysis

Logistic regression		Number of obs	=	100
		LR chi2 (5)	=	31.16
		Prob > chi2	=	0.0000
Log likelihood = -45.504982		Pseudo R2	=	0.2551

Source: Researchers' Data Analysis

Figure 3
Log Likelihood in Logistic Regression Model

The model showed that the log likelihood is greater than the initial logistic regression. However, through trial and error and by using the forward step method, the most significant variable was added first, and by using the backward stepwise method, variables were deleted that were not significant. This showed that the log likelihood is around -45; therefore, an appropriate log likelihood is around -45, which matches the model construction and significance (*p-value<0.001*). The equation can then be constructed as follows:

$$log\left(\frac{p}{1-p}\right) = 11.556 - 0.645\ (A) - 1.570\ (B) - 2.139\ (C) - 0.739\ (D) - 0.984\ (E) \quad (3)$$

Where,
A : the unit changes in the Employment variable
D : the unit changes in the Followers variable
B : the unit changes in the Tenor variable
E : the unit changes in the Following variable
C : the unit changes in the Posting at Midnight variable

The coefficient for constant and independent variables was used later for scoring calculations. The coefficient for each independent variable was negative and represented a negative relationship between the independent variables; this means that if the independent variable is increased to a higher category (higher risk), the probability of defaulting would be reduced. Here are detailed interpretations for each of the independent variables:
- Every change to the unit in the Employment variable will reduce the log of default compared to non-default by 0.645 times
- Every change to the unit in the Tenor variable will reduce the log of default compared to non-default by 1.570 times
- Every change to the unit in the Posting at Midnight variable will reduce the log of default compared to non-default by 2.139 times
- Every change to the unit in the Followers variable will reduce the log of default compared to non-default by 0.739 times
- Every change to the unit in the Following variable will reduce the log of default compared to non-default by 0.984 times

The next statistical process was the logistic regression for all categories in each variable. The results of the logistic regression for each category are exhibited in Table 9:

Table.9
Logistic Regression Model with Category

		B	Std. Err.	z	P> |z|
Employment	Wiraswasta	-1.089	0.964	-1.13	0.258
	Karyawan	-1.467	0.609	-2.41	0.016
Tenor	1-6 Bulan	-1.847	0.605	-3.05	0.002
Posting in Midnight	0-3 Kali	-2.510	1.139	-2.20	0.028
Followers	400-519 Followers	-0.209	0.818	-0.26	0.798
	520-899 Followers	-0.672	0.845	-0.79	0.427
	>=900 Followers	-2.454	0.916	-2.68	0.007
Following	600-899 Following	-1.369	0.922	-1.48	0.138
	300-599 Following	-1.796	0.956	-1.88	0.060
	0-299 Following	-3.256	1.145	-2.92	0.003
	Constant	6.007	1.842	3.26	0.001

Source: Researchers' Data Analysis

4.6. Scoring Calculation

To calculate the score for each category, initially, the scaling format had to be set. First, bad borrower probability was considered to be 3:1 within the total sample: the total number of default borrowers was 100:31, which was rounded to 3:1; alternatively, through bad/good odds, it can be seen as 69/31, which can also be rounded to 3:1. So, the odds involved in the scoring calculation is 3. Furthermore, the baseline score and double odds (*pdo*) are set as 500 and 20, respectively (Siddiqi, 2005); this is because these are the baseline score and point to commonly used double odds. Next, calculate factor and offset value. The parameter results are as follows (see Table 10):

Table 10
Parameter Value

Parameter	Value
Odds	3
Pdo	20
Score	500
Factor	$\frac{20}{\ln(2)} = 28.8539$
Offset	$500-(28.8539 \times \ln(5)) = 468$

Source: Researchers Data Analysis

After determining the parameter for the scoring calculation, the score for each category was calculated. The results are as follows (see Table 11):

Table 11
Scoring Calculation

VARIABLE	CODING	GROUP CATEGORY	WOE	COEFFICIENT	SCORE
	1	Other	-0.667		15
EMPLOYMENT	2	Wiraswasta	0.298	-0.645	33
	3	Karyawan	0.345		33
TENOR	1	7–12 Months	-0.746	-1.570	-7
	2	1–6 Months	0.547		52
POSTING_ MIDNIGHT	1	4–8 Times	-1.493	-2.139	-65
	2	0–3 Times	0.109		34
	1	0–399 Followers	-0.289	-0.739	21
FOLLOWERS	2	400–519 Followers	-0.289		21
	3	520–899 Followers	0.087		29
	4	>=900 Followers	0.499	-0.739	38
	1	>=900 Following	-0.512	-0.984	12
FOLLOWING	2	600–899 Following	-0.330		18
	3	300–599 Following	-0.012		27
	4	0–299 Following	0.726		48

Source: Researchers' Data Analysis

Table 11 indicates that the score increased from the highest risk category to the lowest risk category; the first category had the lowest score, while the first category had the highest score. Therefore, it could be concluded that the reference category matches the assumption; this is because it has the lowest score, which represents the highest risk.

4.7. Cross-Validation Test

The datasets were validated using a cross-validation test. Firstly, the datasets were separated into training and validation samples. This paper used 100% of the data for training and 50% data for validation due to the limited data referred to by Siddiqi (2005). After that, the cross-validation test was initially conducted using a default cut-off value of 0.5 followed by another cut-off value (see Table 12):

Table 12
Cross Validation Test Using Cut-Off Value 0.5

Observed		Predicted		Corrected Rate
		Non-Default	Default	
Training	Non-default	64	6	91.4%
	Default	12	18	60.0%
Hit Rate				82.0%
Misclassification Rate				18.0%
Validation	Non-default	25	2	86.2%
	Default	6	15	71.4%
Hit Rate				80.0%
Misclassification Rate				20.0%

Source: Researchers' Data Analysis

Table 12 shows that the training accuracy rate of discriminating the non-default borrowers was 91.4%, the training sensitivity rate was 60.0%, and the overall hit rate during training sample was 82.0%. Meanwhile, in the validation sample, the accuracy rate declined at 86.2%. However, the sensitivity rate increased to 71.4%, while the hit misclassification rate decreased to 80.0%. This shows that the accuracy rate is higher than the sensitivity rate.

Using a default cut-off value of 0.5, the discriminant capability for default borrowers would be lower than the non-default borrowers; this would result in non-equilibrium discrimination for both types of borrowers (Tsai, 2009). Therefore, to upgrade the discriminant capability of default borrowers, the optimum cut-off value should be represented as a sensitivity rate. The optimum cut-off value is determined by plotting the hit rate, sensitivity rate, and accuracy rate; the cut-off value ranges from 0.5-0.95. Based on the cut-off value plot, the optimum cut-off value is 0.3, where the curve of hit rate, correct percentage of default borrowers, and correct percentage of non-default borrowers are intersected (see Figure 4). Table 13 depicts the classifications, with a cut-off value of 0.3:

Table 13
Cross-Validation Test Using a Cut-Off Value of 0.3

Observed		Predicted		Corrected Rate
		Non-Default	Default	
Training Sample	Non-default	56	14	80.0%
	Default	6	24	80.0%
Hit Rate				80.0%
Misclassification Rate				20.0%
Validation Sample	Non-default	17	12	58.6%
	Default	2	19	81.0%
Hit Rate				68.0%
Misclassification Rate				32.0%

Source: Researchers Data Analysis

According to Table 12, the training hit rate, sensitivity rate, accuracy rate had the same rate of 80.0% with a misclassification rate of 20.0%. Meanwhile, for the validation sample, the hit rate is 68.0%, the sensitivity rate is 81.0%, the accuracy rate is 58.6%, and the misclassification rate is 32.0%. This shows that there is a declining hit and accuracy rate when 0.3 is used as the cut-off value. However, the sensitivity rate has increased for both the training and validation samples; this is good because our research focuses on predicting default borrower probability. It can be concluded that when using a cut-off value of 0.3, the model can upgrade the discriminant capability of default borrowers.
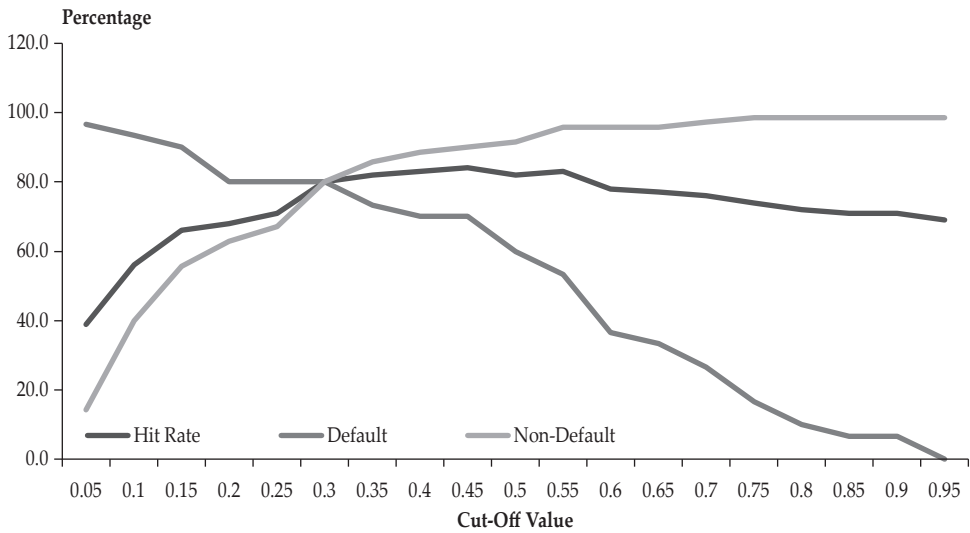
Source: Researchers' Data Analysis

Figure 4
Fitted Cut-Off Value of 0.3

4.2. Robustness Test

Two scenarios were used for the robustness test. First, the model was tested using demographic and historical payment variables. It was then tested using only social media variables. The number of datasets is the same as the initial model development; 100 samples of data using logistic regression.

The first scenario showed the same result; only Employment and Tenor variables were statistically significant ($p < 0.05$). The results are as follows:

Table 14
Robustness Test using Demographic and Historical Payment Variables

Variable	Coef.	P>z
Gender	0.21	0.689
Marital status	-0.68	0.259
District	-0.3	0.363
Income	-0.02	0.939
Employment	-0.72	0.013
Instalment	0.27	0.15
Tenor	-1.62	0.003
Cons.	3.5	0.102

Source: Researchers' Data Analysis

Meanwhile, the second scenario showed the same results; Posting Time, Followers, and Following are significant ($p < 0.05$), in addition to the number of posts per month, which was also significant. The results are as follows:

Table 15
Robustness Test using Social Media Variables

Variable	Coef.	P>z
IG Month	-0.22	0.458
Posting Midnight	-2.12	0.038
Religion Account	-0.62	0.136
Followers	-0.82	0.007
Following	-1.1	0.001
IG Post -Month	-1.4	0.009
Cons.	11.68	0.00

Source: Researchers Data Analysis

4.3. Analysis

Compared to previous studies, the increment rate is better than that of the model of Chen et al. (2016) with an increment rate of 3.16%, but the model used by Tan and Phan is better with an increment rate of 18%. The differences in the AUC could be driven by the number of variables and samples used in each study. Tan and Phan (2016) used 1,047 borrowers with nine independent variables categorised into demographics, preferences, and network variables. Meanwhile, Chen et al. (2016) used a large number of datasets from 30,337 borrowers with 36 dependent variables categorised as demographics, tweets, and networks. Therefore, for future research, it would be better to increase the sample number and involve more variables; this would improve accuracy and achieve better results in determining the predictability level and increasing the model's development.

As well as the differences in the increments of correction rates, social media was proven to improve credit scoring assessments according to the AUC and misclassification. In this paper, it was discovered that the independent variables Posting Time, Followers, and Following from social media datasets fulfil all of the requirements for involvement in company credit scoring, such as significance levels and predictive power. The Tenor and Employment variables were also appropriate variables for use in the scorecard model development. Social media variables that significantly impact on credit scoring development is in line with Chen et al.'s (2016) study, which determined that Posting Time, Followers and Following were statistically significant.

According to the comparison between the findings from this and previous studies, the company could consider posting time, number of followers and number of following to their social media assessment. It is clear that people who tend to post at midnight, have fewer followers and tend to follow large number of people on social media tend to default. However, Chen et al.'s (2016) results contradict this as they regard those following a smaller number of people as those that will tend to default on a loan. This difference might be affected by the number of samples used, the type of social media being observed, and the demographic characteristics of the sample. Therefore, a larger number of samples with various characteristics would be important to increase the accuracy of the results.

V. CONCLUSION AND RECOMMENDATION

The objectives of this research were to identify social media variables that could be used to support an Islamic online P2P lending company's credit scoring model and determine the predictability levels of credit scoring through the use of social media data within an Islamic online P2P lending company. Through the research, a conclusion and recommendations have been determined.

5.1. Conclusion

Using the data of 100 borrowers during the period March 2017–January 2019 and through the analysis of credit scorecard and logistic regression, it can be concluded that:

1. Posting at Midnight, Followers, and Following are social media variables that could be predictors of default probability because those three variables fulfilled the goodness of fit criteria, which comprises Weight of Evidence (WoE), Information Value, and p-value. Furthermore, Tenor and Employment are also included as variables that could be used as predictors.

2. According to the misclassification rate, adding social media variables to the model could minimise misclassification rates. The model improvement was proven through Area Under Curve (AUC), which showed that by combining selected variables with demographics, historical payments, and social media data, the AUC increases by 6.6% compared to models that only use the demographics and historical payment variables; this indicates that the rate of successful classification in the predicting model could be improved.

5.2. Recommendation

According to the conclusion, the researchers derive the recommendations as follows:

1. The company could develop their social media assessments by considering Posting Time, Followers, and Following as well as the Tenor and Employment variables. Furthermore, the company should be more cautious with people that frequently post at midnight, have fewer followers, are given credit durations of longer than six months, is employed with a low salary, and who follow a large number of people on their social media platform, especially on Instagram. Applicants with those criteria should be given a low score for their credit rating to minimise future default risk.

2. For future research, it would be better to use data from a longer period; here, the data was limited because the company was only built in 2017. By taking data from a longer period and from a larger population, the model would be more accurate. Furthermore, it is effective to use more independent variables, such as the sentiment of a borrower's status on social media, to obtain a specific and more detailed result. The collection methods for social media data could be improved by using more efficient collection tools.

3. For the regulator and supervisor (Bank Indonesia and Otoritas Jasa Keuangan), social media data might be considered to develop a credit scoring model. However, the user privacy aspect must be considered if using social media

data for assessments related to the regulations from Undang-Undang No. 11 Tahun 2008 article 26 and Peraturan Pemerintah No. 82 Tahun 2012 , which state that it is obligatory to have consent and agreement from the related individual to use their online social media data. For this study, the borrowers previously agreed for their social media data to be assessed by the company.

REFFERENCES

Ancri, C. (2016, October 19). *Fintech Innovation : An Overview.* Retrieved November 30, 2018, from World Bank Publication: http://pubdocs.worldbank.org/en/767751477065124612/11-Fintech.pdf

Asosiasi Penyelenggara Jasa Internet Indonesia. (2018). *Laporan Survei Penetrasi & Profil Perilaku Pengguna Internet Indonesia.* Retrieved June 10, 2019, from https://apjii.or.id/survei

Badan Pusat Statistik. (2017). *Produk Domestik Regional Bruto Kabupaten/Kota di Indonesia 2013-2017.* Retrieved April 18, 2019, from https://www.bps.go.id/

Bagaskara, B. P., & Setiawan, C. (2016). Non-Performing Financing (NPF) and Cost Efficiency of Islamic. *Journal of Emerging Issues in Economics, Finance and Banking (JEIEFB), 5*(1), - Retrieved September 29, 2018, from http://globalbizresearch.org/economics/images/files/74631_T632_JEIEFB_Chandra%20Setiawan_Bhirawa%20Praditya%20Bagaskara.pdf

Bank Indonesia. (n.d.). Retrieved October 08, 2018, from Bank Indonesia Official Website: https://www.bi.go.id/id/edukasi-perlindungan-konsumen/edukasi/produk-dan-jasa-sp/fintech/Pages/default.aspx

Bank Indonesia. (2015). *Profil Bisnis UMKM.* Retrieved November 25, 2018, from https://www.bi.go.id/id/umkm/penelitian/nasional/kajian/Pages/Profil-Bisnis-UMKM.aspx

Bank Indonesia. (2018). *Survei Perbankan Triwulan III - 2018.* Retrieved November 25, 2018.

Basel. (2000). Principles for The Management of Credit Risk. 1. Retrieved December 05, 2018, from https://www.bis.org/publ/bcbs54.pdf

Bolton. (2009). Logistic Regeression and Its Application in Credit Scoring. Retrieved December 01, 2018, from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.1024.2660&rep=rep1&type=pdf

Brown, K., & Peter, M. (2014). *Credit Risk Management.* Great Britain. Retrieved November 03 , 2018, from https://www.ebsglobal.net/documents/course-tasters/english/pdf/h17cr-bk-taster.pdf

Chen, F., Guan, C., Guo, G., Liu, Q., Wu, L., & Zhu, F. (2016). From Footprint to Evidence: An Exploratory Study of Mining Social Data for Credit Scoring. *ACM Transactions on the Web (TWEB).* Retrieved November 15, 2018, from https://dl.acm.org/citation.cfm?id=2996465

Chen, L., Li, Q., & Zeng, Y. (2018). The mechanism and effectiveness of credit scoring of P2P lending platform: Evidence from Renrendai.com. *China Finance Review International.* Retrieved November 03, 2018, from https://www.emeraldinsight.com/doi/abs/10.1108/CFRI-06-2017-0156

Claessens, S., Frost, J., Turner, G., & Zhu, F. (2018, September). BIS Quarterly Review, September 2018 29 Fintech Credit Markets around The World: Size,

Drivers and Policy Issues. *BIS Quarterly Review*, p. 33. Retrieved June 11, 2018, from https://www.bis.org/publ/qtrpdf/r_qt1809e.pdf

Cortina, J. J. (2018, April). *The Fintech Revolution: A Threat to Global Banking?* Retrieved June 11, 2019, from World Bank Document: http://documents. worldbank.org/curated/en/516561523035869085/pdf/125038-REVISED-A-Threat-to-Global-Banking-6-April-2018.pdf

Dellarocas, C., Wei, Y., Yildirim, P., & Van den Bulte, C. (2016). Credit Scoring with Social Network Data. *Marketing Science*. Retrieved November 03, 2018, from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2475265

Emekter, R., Jirasakuldech, B., Lu, M., & Tu, Y. (2014). Evaluating Credit Risk and Loan Performance in Online Peer-to-Peer (P2P) lending. *Journal Applied Economics*. Retrieved November 03, 2018, from https://www.tandfonline.com/ doi/abs/10.1080/00036846.2014.962222

Fatemi, A., & Fooladi, I. (2006). Credit Risk Management: A Survey of Practices. *Managerial Finance*, 227-233. Retrieved November 03, 2018, from https://www. emeraldinsight.com/doi/abs/10.1108/03074350610646735

Fight, A. (2004). *Credit Risk Management.* Oxford: Elsevier Butterworth-Heinemann. Retrieved November 03, 2018, from https://books.google.co.id/books?hl=id&l-r=&id=wd15dswnQWgC&oi=fnd&pg=PP1&dq=credit+risk+manage-ment&ots=J7mHDienpt&sig=FivaBXFmLFDwYT6sUVPZwKHcCs&redir_es-c=y#v=onepage&q=credit%20risk%20management&f=false

Governor, B. I. (2017, Agustus 28). Transaksi fintech diperkirakan US$ 18,65 miliar. (A. A. Mustami, Interviewer, & W. Rahamawati, Editor) Kontan. co.id. Retrieved November 30, 2018, from https://nasional.kontan.co.id/news/ transaksi-fintech-diperkirakan-us-1865-miliar

Greuning, H., & Iqbal, Z. (n.d.). Risk Analysis for Islamic banks (English). Retrieved December 04, 2018, from http://documents.worldbank.org/curated/ en/688471468143973824/Risk-analysis-for-Islamic-banks

Gujarati, D. (2004). *Basic Econometrics.* Retrieved May 25 , 2019, from http://www. afriheritage.org/TTT/2%20Basic%20Econometrics%20-%20Gujarati[1].pdf

Hsieh, & Lee. (2011). Data Mining in Building Behavioral Scoring Model. Retrieved November 14, 2018, from https://www.researchgate.net/ publication/224208444_Data_Mining_in_Building_Behavioral_Scoring_ Models

Kementerian Keuangan. 2008. Undang - Undang No. 11 Tahun 2008. Retrieved 6 December, 2019 from http://www.jdih.kemenkeu.go.id/ fullText/2008/11TAHUN2008UU.htm

Kementerian Komunikasi dan Informatika. 2012. Peraturan Pemerintah No. 82 Tahun 2012. Retrieved December 6, 2019 from https://jdih.kominfo.go.id/produk_hukum/view/id/6/t/ peraturan+pemerintah+republik+indonesia+nomor+82+tahun+2012

Kocenda, & Vojtek. (2006). Credit Scoring Methods. Retrieved November 14, 2018, from https://pdfs.semanticscholar.org/fe0c/ ac08a8a5206501b95fe73bac043ab5ac8ddb.pdf.

Kosinskia, M., Stillwella, D., & Graepelb, T. (2013). Private Traits and Attributes are Predictable from. *Proceedings of the National Academy of Sciences of the United States of America*. Retrieved June 25, 2019, from http://discovery.ucl. ac.uk/1395203/

Marifatika, E. (2015). *Pembagian Waktu AM dan PM*. Retrieved March 01 , 2019, from kelasbahasainggris.com: http://kelasbahasainggris.com/pembagian-waktu-am-dan-pm/

Midi, H., Rana, S., & K., S. S. (2013). Collinearity Diagnostics of Binary Logistic Regression Model. *Journal of Interdisciplinary Mathematics*. Retrieved May 31, 2019, from https://www.tandfonline.com/doi/abs/10.1080/09720502.2010.1070 0699

Novalia, I. (2018, Agustus). Mengenal Investasi P2P Lending Syariah dan Cari Tahu Cara Investasinya. Retrieved June 25, 2019, from https://www.finansialku.com/p2p-lending-syariah/

Nyong'o, & N., J. (2009). *The Relationship between Credit Risk Management and Non-Performing Loans in Commercial Banks in Kenya.* Retrieved November 03, 2018, from http://erepository.uonbi.ac.ke/bitstream/handle/11295/76217/Nyong%60oThe%20relationship%20between%20credit%20risk%20management%20and%20non-performing%20loans%20in%20commercial%20banks%20in%20Kenya.pdf?sequence=3

OJK, D. o. (2018, Desember 13). Yang Membuat OJK Tak Berdaya Hadapi Fintech P2P Lending Ilegal. (V. F. Thomas, Interviewer) Tirto.id. Retrieved June 11, 2019, from https://tirto.id/yang-membuat-ojk-tak-berdaya-hadapi-fintech-p2p-lending-ilegal-dbLP

OJK, D. o. (2019, March 01). Kredit Macet Pinjaman Online Makin Tinggi, Apa Jadi Bom Waktu? (D. C. Syafina, Interviewer) Retrieved June 11, 2019, from https://tirto.id/kredit-macet-pinjaman-online-makin-tinggi-apa-jadi-bom-waktu-dhZM

Otoritas Jasa Keuangan & Asosiasi Fintech Indonesia. (2017). Financial Technology (Fintech) di Indonesia. Retrieved June 10, 2019, from http://www.ibs.ac.id/img/doc/MDH%20-%20FinTech%20IBS%20June%202017.pdf

Otoritas Jasa Keuangan. (2016). Peraturan Otoritas Jasa Keuangan Nomor 77/POJK.01/2016. Retrieved November 25, 2018, from https://www.ojk.go.id/id/regulasi/otoritas-jasa-keuangan/peraturan-ojk/Documents/Pages/POJK-Nomor-77-POJK.01-2016/SAL%20-%20POJK%20Fintech.pdf

Otoritas Jasa Keuangan. (2018, September). *Ikhtisar Data Keuangan Fintech (P2P Lending) Periode September 2018.* Retrieved November 04, 2018, from OJK Official Websiter: https://www.ojk.go.id/id/kanal/iknb/data-dan-statistik/fintech/Pages/Ikhtisar-Data-Keuangan-Fintech-(Peer-To-Peer-Lending)-Periode-September-2018.aspx

Otoritas Jasa Keuangan. (2018). *Laporan Profil Industri Perbankan Triwulan III - 2018.* Retrieved November 25, 2018, from https://www.ojk.go.id/id/kanal/perbankan/data-dan-statistik/laporan-profil-industri-perbankan/Documents/LPIP%20TW%20III-18.pdf

Otoritas Jasa Keuangan. (2018, September). *Perkembangan Fintech Lending (Pendanaan Gotong-Royong On-Line).* Retrieved November 03, 2018, from OJK Official Website: https://www.ojk.go.id/id/berita-dan-kegiatan/siaran-pers/Documents/Pages/Siaran-Pers-Satgas-Kembali-Temukan-182-Fintech-Peer-To-Peer-Lending-Tanpa-Izin/180905%20Statistik%20Fintech%20Lending%20Indonesia-in%20Bahasa.pdf

Parveen, H., & Showkat, N. (2017). Non-Probability and Probability Sampling. Retrieved June 25, 2019, from https://www.researchgate.net/publication/319066480_Non-Probability_and_Probability_Sampling/download

PWC. (2016, March). *Blurred lines:How FinTech is shaping Financial Services.* Retrieved September 30, 2018, from https://www.pwc.de/de/newsletter/finanzdienstleistung/assets/insurance-inside-ausgabe-4-maerz-2016.pdf

PWC. (2018). *Indonesia Banking Survey.* Retrieved November 01, 2018, from https://www.pwc.com/id/en/publications/assets/financialservices/2018-indonesia-banking-survey.pdf

Scorto. (n.d.). *Behavioral Scoring*. Retrieved November 03, 2018, from https://scorto.com/behavioral-scoring/

Siddiqi, N. (2005). *Credit Risk Scorecards : Developing and Implementing Intelligent Credit Scoring.* John Wiley & Sons, Inc. Retrieved May 31, 2018, from https://pdfs.semanticscholar.org/dd5c/7f59d20d9a00d4c93e3d6a7e9973f3462e7e.pdf%20[27%20May%202018

Siddiqi, N. (2006). *Credit Risk.* John Wiley & Sons, Inc. Retrieved May 31, 2019, from https://pdfs.semanticscholar.org/dd5c/7f59d20d9a00d4c93e3d6a7e9973f3462e7e.pdf%20[27%20May%202018

Statista. (n.d.). Retrieved October 08, 2018, from https://www.statista.com/outlook/295/120/fintech/indonesia#market-arpu

Tabagari, S. (2015). Credit Scoring by Logistic Regression. Retrieved November 15, 2018, from http://dspace.ut.ee/handle/10062/47572

Tan, & Phan. (2016). Social Media-Driven Credit Scoring: the Predictive Value of Social Structures. Retrieved November 04, 2018, from https://www.semanticscholar.org/paper/Social-Media-Driven-Credit-Scoring%3A-the-Predictive-Tan-Phan/2f1ce382e2be6ff6c70e2a43e0197d89426992c9

Thomas, L. C. (2000). A survey of credit and behavioural scoring: forecasting financialrisk of lending to consumers. *International Journal of Forecasting*. Retrieved December 04, 2018, from http://socsci2.ucsd.edu/~aronatas/project/academic/A%20survey%20of%20credit%20and%20behavioural%20scoring%20Forecasting%20fina.pdf

Todorof, M. (2018). Shariah-compliant FinTech in The Banking Industry. *ERA Forum*.

Tsai, e. a. (2009). The Consumer Loan Default Predicting Model – An Application of DEA–DA. *Expert Systems with Applications*. Retrieved November 03, 2018, from https://www.semanticscholar.org/paper/The-consumer-loan-default-predicting-model-An-of-Tsai-Lin/d75535e84eb7a68b9aefbe7765033d38298b1f63

Wibowo, P. P. (2017, September). *Innovation & Fintech Innovation Development in Retail Payment in Indonesia.* Retrieved October 08, 2018, from https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=9&cad=rja&uact=8&ved=2ahUKEwjXmbSWzN3iAhXOXisKHa-PAWkQFjAIegQIBxAC&url=https%3A%2F%2Fwww.bi.go.id%2Fid%2Finstitute%2Fkegiatan%2FFlagship%2FContents%2FDay4_Session3_Innovation%2520and%2520Fintec

World Bank. (2017). *Global Findex*. Retrieved June 10, 2019, from https://globalfindex. worldbank.org/sites/globalfindex/files/chapters/2017%20Findex%20full%20 report_chapter2.pdf

Yap, B. W., Ong, S. H., & Husain, N. (2011). Using Data Mining to Improve Assessment of Credit Worthiness via. *Expert Systems with Applications*. Retrieved November 14, 2018, from https://www.sciencedirect.com/science/ article/pii/S0957417411006749