Jurnal Pendidikan dan Konseling
Volume 4 Nomor 5 Tahun 2022
E-ISSN: 2685-936X dan P-ISSN: 2685-9351
Universitas Pahlawan Tuanku Tambusai

A New Credit Scoring Model For Vehicle Leasing Company

Jonathan Tandarma Tjhin[1], Gunawan Wang[2]
[1,2] Master Management System Information, Binus University,
Department of Computer Science, Indonesia
Email : jonathan.tjhin@binus.ac.id, gwang@binus.edu

Abstrak

Usaha kecil dan menengah menjadi salah satu bisnis yang terdampak akibat penyebaran virus corona. Situasi pandemi di Indonesia menyebabkan penderitaan besar pada perusahaan-perusahaan ini. Untuk mencegah kerugian di masa pandemi saat ini. PT XYZ memutuskan untuk membuat model penilaian kredit untuk memprediksi risiko dari calon pelanggan mereka. Model akan terdiri dari dua jenis. Yang pertama adalah penilaian atau kartu skor sistem pakar. Data yang diperoleh dari sistem pakar nantinya akan dimasukkan ke dalam machine learning menggunakan metode statistik untuk mendapatkan model credit scoring. Kerangka kerja CRISP-DM akan digunakan untuk memandu proses pembuatan untuk memastikan keluaran model yang andal.

Kata Kunci: *Credit Scoring, Credit Scorecard, Model Statistik, Usaha Kecil Menengah, SVM, Deep Learning, CRISP-DM*

Abstract

Small and medium enterprises are one of the affected business due to the spread of the coronavirus. The pandemic situation in Indonesia cause a great suffer on these enterprises. To prevent loss during the pandemic time right now. PT XYZ decide to create a credit scoring model to predict the risk from their prospective customer. The model will consist of two types. The first is the judgmental or expert system scorecard. The data obtained from the expert system later will be fed into machine learning using statistics method to obtain a credit scoring model. The CRISP-DM framework will be used to guide the creation process to ensure a reliable model output.

Keywords: *Credit Scoring, Credit Scorecard, Statistics Model, Small And Medium Enterprises, SVM, Deep Learning, CRISP-DM*

INTRODUCTION

In the late January 2020, the World Health Organization (WHO) has announced the Public Health Emergency of International Concern (PHEIC) in accordance with the outbreak of novel coronavirus / COVID 19 virus. Since then, the world had been fighting to stop the spread of the new deadly corona virus [1]. Furthermore, the virus also has an impact on global economic. According to the report from World Bank, all country in all region had been impacted to the degree that the economic growth is contracted [2]. Indonesia is one of the country that is affected by the pandemic situation, fortunately the economic still has a positive growth rate. Nevertheless, this figures are dropped by a large margin compared with the last year statistics at the same time [3].

This economic impacting not only big enterprises but many of the SMEs (Small and Medium Enterprise) are facing the impact of the pandemic. Unlike monetary crisis in 1998, when the Indonesian

SMEs become the hero for national economic growth. In 2020 pandemic, there are a lot of government regulation that restrict people mobility, office work time, public places, and many more, causing the SMEs facing a great loss [4]. Take for example, SMEs that in the field of food services such as restaurants and food stalls suffer the most due to the government regulations. But not all SMEs are facing a great despair in this situation, a few of SMEs still growing for example logistics, laundries, and telecommunication service provider has seen the increase since the pandemic breakout [5]. PT XYZ is one of the SMEs that provide rental services for their customers. PT XYZ provides electric motorcycle and tricycle as their main business. According to the report by [6], the rental business facing a great contraction in early 2020, but has been steadily moving toward a good trend. To avoid losses during this pandemic event, PT. XYZ wants to create a credit scoring model to assess their customer.

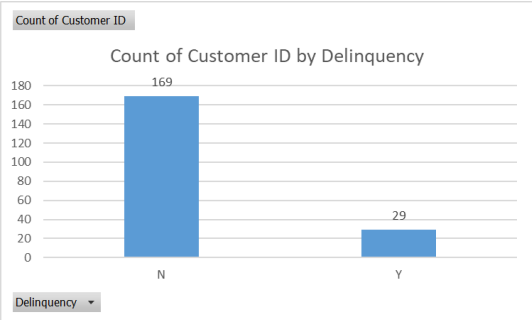

Figure 1: Customer vs Delinquency PT XYZ as of May 2021

From the data that has been collected, as of figure 1, shows that from the total of 198 customers, there are 29 which turn out to be delinquent (fail to pay the rent service on time or did not pay at all). So we can conclude that the NPL rate from PT XYZ around 14%. Quite a staggering amount compare to the National NPL ration which is only 3.1% in Feb 2021. [7] Credit scoring model can prove to be a good tool to assess prospective customer, so that the company has a better understanding on the risk before making a certain contract with the prospective customer. [8] This paper will discuss on the credit scoring model and expert scorecard that will be deployed on the PT XYZ use case. The methodology used will be the CRISP-DM (Cross Industry Standard Process – Data Mining). The model will be used to measure the risk of prospective customer, thus empowered the organization to decide whether to accept or reject the applicant of prospective customer.

Credit Scoring

There are many explanations coined by various previous researcher. In 1997, the terms credit scoring defined by [9] as process to determine the delinquency probability from prospective customer using statistical model such as credit scorecard. Later on, in 2011, [10] described credit scoring as a statistic tools used to assist finance company, in particular banking company, to be able to determine the risk of a prospective customer to go delinquent or not.

Overall all the researchers agree that credit scoring is a statistics technic to transform customer data into a set of ranking that can be used as a reference to give certain amount of credit to prospective customer [8], [11].

A good credit score has a great link to the success of a finance company. According to [12], a good credit scoring model can help a financial company to succeed in the early step of fundraising during the startup process.

Credit Scorecard

Credit scorecard is a terminology to explain a statistical model to estimate credit worthiness from prospective customer based on a group of variable [13].

There are 3 types of credit scorecard, there are judgmental, statistical, and hybrid. Judgmental or expert system is a type of scorecard that created based on expert judgement inside an organization. While statistical is based on historical data. On the other hand, hybrid type is a combination between expert system and statistical. A model is created with statistical data, later will be weighted by expert system [14].

Critical Factor to Assess Credit Worthiness

According to [15], there are 5 critical factor to assess credit worthiness, there are Character, Capacity(the capability to pay loan), Capital, Condition of Business, and Collateral. These 5C are the most common factor used to assess credit worthiness. The research conducted by [16] on Financial Institution based on Indonesia, add another C which is Cash Flow. The Cash Flow factor measure how well the condition of candidate money management.

However due to the nature of the business model of PT. XYZ, it is not possible to collect Capital characteristic. Since PT. XYZ does not stands as a financial company rather as a rental company. Thus we will drop the Capital characteristic. The paper only takes 5C as the main critical factor to assess candidate credit worthiness. There are, Character, Capacity, Condition of Business, and Collateral.

CRISP-DM

CRISP-DM stands for Cross Industry Standard Process for Data Mining. It is a framework created to make large data mining projects more structured, less costly, more reliable and repeatable, and faster [17].

This framework consists of several phases as follow:

a. Business Understanding

The initial phase focused on gathering and understanding the requirement and objectives of the project.

b. Data Understanding

Initial data collection, data identification to gather information and insights from the data.

c. Data Preparation

Perform all necessary activities to the raw data into clean data that will be used for calculation or modelling with available tools.

1. Modelling

   Various modeling techniques are selected and applied.
2. Evaluation

   Evaluate the model from previous step. The evaluation process is to determine whether the result from the model can achieved business requirement.
3. Deployment

   Deploying the model in customer environment. Deploying the model can range from a simple reporting task or as complex as implementing a repeatable data mining process.

METHOD

Business Understanding

PT XYZ is one of the SMEs which provide rental services for their customer. In concern due to the pandemic spread, PT XYZ want to enhance their applicant process with credit scoring. It is known that credit scoring can help the company to have a better risk management [9].

PT XYZ has a team specialist in managing the applicant form. We will focus on this team, to create the judgmental (expert system) credit scorecard. Using their experience, we will create a survey later will be feed to the machine learning to create a model for PT XYZ customer scoring.

In this step we also defined what is the meaning of delinquent and the group of credit worthiness. It is decided by FGD with the expert staff of PT XYZ, that delinquent is the condition of a customer whose did not pay or miss the due date by 1 day. The target group of credit worthiness are decided into 3 categories, there are Excellent – Worth Giving, Good – Give credit with consideration, Poor – Not Worth Giving.

Data Understanding

The 5C factors (Character, Capacity, Condition of Business, Cash Flow, and Collateral) will be used as the basis for the data. To asses this factor, we need to break them into smaller category. The Character factor will be assessed by examining the behavior of the prospective candidate. This has been formulated into 6 variables namely character, motivation, attitudes, family-relation, social-relation, and religiosity. For the complete question – answer and score sheet see appendix 1.

The Capacity factor will be assessed by examining the business that the prospective candidate owns. We will examine the profit from the current or planned business, the accounts receivable, and the expected or current performance of the business. For complete question-answer and score sheet see appendix 2. The Cash Flow performance will be used as a parameter to measure the cash flow factor. See appendix 3 for complete question-answer and score sheet.

Condition of business factor will be measured by assessing the business prospect, sector, and the product type that they offer. Also we will add business rivalry condition to take into account the condition of local competition for each prospective candidate. See appendix 4 for complete question-answer and score sheet. The last factor is Collateral. Collateral factor will be assessed by examining on the liquidity and the collateral ratio to loan. It is important to know the collateral that can be used by the prospective candidate. As it is a safeguard if the prospective candidate goes delinquent. See appendix 5 for complete question-answer and score sheet.

The data for this research can be seen as the figure below.

Table 1: Variables of the Critical Factors

No	Data	Type
1	Character	Polynominal
2	Motivation	Polynominal
3	Attitudes	Polynominal
4	Family –Relation	Polynominal
5	Social – Relation	Polynominal
6	Religiosity	Polynominal
7	Profit From Business	Polynominal
8	Accounts Receivable	Polynominal
9	Business Performance	Polynominal
10	Cash Flow Performance	Polynominal
11	Business Prospect	Polynominal
12	Business Sector	Polynominal
13	Business Product Type	Polynominal
14	Business Rivalry	Polynominal
15	Collateral Liquidity	Polynominal
16	Collateral to Loan ratio	Polynominal
17	Delinquent	Boolean

Data Preparation

The survey collected from prospective customer is in the form of paperwork. This paperwork then translated into excel worksheet. Several preparations done through the data are as follow:

a. Trim missing value out
b. Perform scoring for each question
c. Determine the credit worthiness of each candidat

1. Modelling

   We will utilize the auto-model feature provided by Rapid Miner software to enable us quickly get a suitable model based on the data we provided. In this study we will focused on the popular predictive model such as logistic regression, deep learning, and SVM.

2. Evaluation

   The performance evaluation will use split-validation with multi hold out validation. Rapid Miner will divided the data into 60:40 (Train:Test). The model will be trained with 60% of data. The remaining 40% of data will be split into 7 subsets. Once the model has finished training, the model will be used to predict on the 7 subset independently. The result from the prediction then will be averaged. The averaged value is the performance of the model.

3. Deployment

   The credit scoring model using statistic method will be presented to the organization. The model then will be discussed to match the company arrangement. The credit scorecard will be refined based on the result of the model from Rapid Miner.

RESULT AND DISCUSSION

Expert Judgmental Scorecard

To collect data, we first create a series of questionnaire to make the expert judgmental scorecard. This scorecard will act as a foundation for our data. The scorecard is divided into 5 main section. Namely Character, Capacity, Cash Flow, Condition of Business, and Collateral. Each section has a series of question that will be scored. The combined scored from all the section will be categorized into 3 categories which is Excellent – Worth Giving, Good – Give credit with consideration, Poor – Not Worth Giving.

The complete question-answer and score will be presented on appendixes.

In order to determine the weight for each section, we conducted a focus group discussion with stakeholder and expert staff from PT. XYZ. The main topic is to determine each of section weight based on the experience and situation from the company. The result of the FGD are as follow:

Table 2: Weight From FGD Result

Criteria	Weight
Character	25
Capacity	20
Cash Flow	20
Condition Of Business	20
Collateral	15
Total	**100**

To calculate the score, we will use simple product formula, where we will calculate the score for each section with the sum of the score for each question multiply by the weight of each section.

Then the overall score will be calculated and categorized into 3 types as follows.

Table 3: Credit Worthiness Category

Score	Category
> 1.000	Excellent
500 - 1000	Good
<500	Bad

We have collected a total 198 samples; which 25 samples are invalid due to missing values. From this sample the result is as figure shown below.

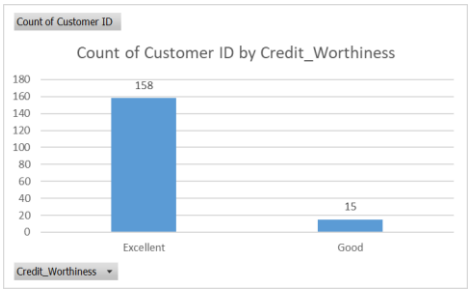

Figure 2: Result Of Scorecard

There are 158 with Excellent category and 15 with Good category. From this survey there are no customer falls under the Bad category.

Credit Scoring Modelling with Rapid Miner

We use Rapid Miner software to help us determining the best model. Rapid Miner has a feature called auto-model. This feature will help us determine the best model for prediction problem. In this research we will focused on logistic regression model, deep learning, and SVM Model.

We choose this model as it is shown as common statistical method to perform predictive analysis [18]. It is good to be noted that there are no perfect model that can be applied to all circumstances. Hence we need to compare different model to find a suitable model based on the data provided [19].

We have collected 198 surveys as of May 2021. Then we perform data pre-processing. From 198 surveys, there are 25 surveys data that cannot be processed due to major missing value in the surveys (more than 5 questions left out blank). Thus only 173 survey data can be processed to make the model.

The result of the auto-model is presented in figure 2 as follow.

Figure 2: Auto-Model Results

Model	Classification Error	Standard Deviation	Gains	Total Time
Logistic Regression	18.0%	±8.4%	0	10 s
Deep Learning	18.0%	±8.4%	-4	22 s
Support Vector Machine	26.0%	±5.5%	-12	25 s

Since the logistic regression model gives the best performance in terms of accuracy and time needed, we will discuss more on the model.

The weight of each variable are shown in table 2 as follow.

Table 3: Weight of Each Variables

Attribute	Weight
Collateral to Loan ratio	0.208127
Social – Relation	0.148826
Cash Flow Performance	0.143306
Business Rivalry	0.100046
Family –Relation	0.092516
Profit From Business	0.084804
Character	0.081591
Business Sector	0.076033
Business Product Type	0.053899
Motivation	0.049163
Accounts Receivable	0.040347
Religiosity	0.03863
Business Prospect	0.031021
Business Performance	0.027586
Collateral Liquidity	0.011386
Attitudes	0.009485

From table 2 we can see that Character, Profit from business, and Family-Relation has bigger weight value compare to other variables. This means that those variable play important role in determining the probability of delinquency from prospective customer.

Table 4: Performance Results

Criterion	Value	Standard Deviation
Accuracy	0.82	0.083666
Classification_error	0.18	0.083666

CONCLUSION

We have discussed on the importance of credit scoring to lower the NPL rate in an organization. By using the credit scorecard in this case the expert judgmental scorecard, we can collect database for creating a predictive model using software such as Rapid Miner. The predictive model can be used to predict the probability of prospective customer to go delinquent or not.

With the recent data collected until May 2021, we get a result of 82% model accuracy using Logistic Regression model. This shows a promising start to develop more complex model.

Future research should focus on how to improve the logistic model and the scorecard. Using ensemble model which combine logistic regression with other predictive model can be explored.

Furthermore, the expert judgmental scorecard should always be updated according to the needs of the company. Especially adjusting the weight of each section may affect the result of each dataset. Exploring a good balance of weight according to the latest condition may be the best option in order to keep up with the latest issue.

REFRENCES

"Coronavirus disease (COVID-19) outbreak." https://www.who.int/westernpacific/emergencies/covid-19 (accessed Apr. 25, 2020).

The World Bank, "The Global Economic Outlook During the COVID-19 Pandemic: A Changed World," The World Bank. 2020, Accessed: 14-Aug-2020. [Online]. Available: https://www.worldbank.org/en/news/feature/2020/06/08/the-global-economic-outlook-during-the-covid-19-pandemic-a-changed-world.

N. Wira Sakti, "Perekonomian Indonesia Pasca-Pandemi Covid-19 Halaman all - Kompas.com," Kompas.com, 2020.

S. Simanjorang, "Nasib UMKM di Tengah Pandemi Covid-19," Investor Daily Indonesia, 2020. https://investor.id/opinion/nasib-umkm-di-tengah-pandemi-covid19 (accessed Aug. 14, 2020).

E. Akbar, "Sektor UMKM yang Meroket dan Tertekan Selama Pandemi Covid-19," Bisnis.com, 2020. https://ekonomi.bisnis.com/read/20200611/12/1251486/ini-sektor-umkm-yang-meroket-dan-tertekan-selama-pandemi-covid-19 (accessed Aug. 14, 2020).

I. Puspitasari, "Bisnis emiten rental kendaraan makin cerah, simak rekomendasi analis berikut," Jakarta, Oct- .

CEIC, "Indonesia Non Performing Loans Ratio, 2003 – 2021 Data," 2021. https://www.ceicdata.com/en/indicator/indonesia/non-performing-loans-ratio (accessed May 26, 2021).

D. J. Hand, S. Y. Sohn, and Y. Kim, "Optimal bipartite scorecards," Expert Syst. Appl., vol. 29, no. 3, pp. 684–690, 2005, doi: 10.1016/j.eswa.2005.04.032.

D. J. Hand and W. E. Henley, "Statistical classification methods in consumer credit scoring: A review," J. R. Stat. Soc. Ser. A Stat. Soc., vol. 160, no. 3, pp. 523–541, 1997, doi: 10.1111/j.1467-985X.1997.00078.x.

H. A. Abdou and J. Pointon, "CREDIT SCORING, STATISTICAL TECHNIQUES AND EVALUATION CRITERIA: A REVIEW OF THE LITERATURE," Intell. Syst. Accounting, Financ. Manag., vol. 18, no. 2–3, pp. 59–88, Apr. 2011, doi: 10.1002/isaf.325.

R. A. Eisenbeis, "Problems in applying discriminant analysis in credit scoring models," J. Bank. Financ., vol. 2, no. 3, pp. 205–219, 1978, doi: 10.1016/0378-4266(78)90012-2.

S. Kaewmungkoon, "the Effect of Credit Scoring Providence on Peer-To- Peer Lending Success : Evidence From Business Platform in Thailand," vol. 7, no. 13, pp. 2191–2201, 2020.

X. Feng, Z. Xiao, B. Zhong, Y. Dong, and J. Qiu, "Dynamic weighted ensemble classification for credit scoring using Markov Chain," Appl. Intell., vol. 49, no. 2, pp. 555–568, 2019, doi: 10.1007/s10489-018-1253-8.

D. Caire, "Building Credit Scorecards for Small Business Lending in Developing Markets," Building, no. November, pp. 1–11, 2004.

J. Nguyen, "Approach to Scorecard Model," 2020. https://rstudio-pubs-static.s3.amazonaws.com/441239_e4a46ef90bd5421bb810e95c06f22df8.html (accessed Aug. 13, 2020).

A. Bhinadi, "SKOR KREDIT DAN PROBABILITAS KREDIT BERMASALAH UNTUK PENYALURAN KREDIT DI LEMBAGA KEUANGAN MIKRO TAHUN 2007: STUDI KASUS …," no. January 2009. 2009, doi: 10.13140/RG.2.2.30008.49921.

R. Wirth, "CRISP-DM : Towards a Standard Process Model for Data Mining," Proc. Fourth Int. Conf. Pract. Appl. Knowl. Discov. Data Min., no. 24959, pp. 29–39, 2000.

A. Arora et al., "Comparison of logistic regression, support vector machines, and deep learning classifiers for predicting memory encoding success using human intracranial EEG recordings," J. Neural Eng., vol. 15, no. 6, 2018, doi: 10.1088/1741-2552/aae131.

F. Louzada, A. Ara, and G. B. Fernandes, "Classification methods applied to credit scoring: Systematic review and overall comparison," Surv. Oper. Res. Manag. Sci., vol. 21, no. 2, pp. 117–134, 2016, doi: 10.1016/j.sorms.2016.10.001.