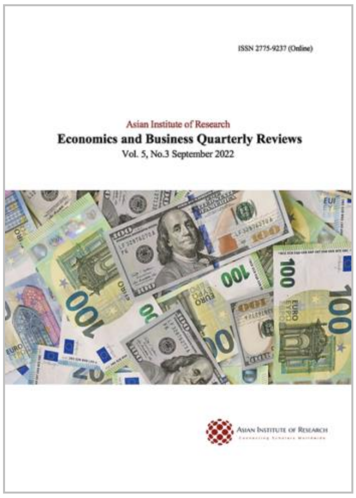

Economics and Business
Quarterly Reviews

Surjaningsih, N., Werdaningtyas, H., Rahman, F., & Falaqh, R. (2022). Predicting Household Resilience Before and During Pandemic with Classifier Algorithms. *Economics and Business Quarterly Reviews*, 5(3), 75-81.

ISSN 2775-9237

DOI: 10.31014/aior.1992.05.03.437

The online version of this article can be found at:
https://www.asianinstituteofresearch.org/

Published by:
The Asian Institute of Research

The Asian Institute of Research
Economics and Business Quarterly Reviews
Vol.5, No.3, 2022: 75-81
ISSN 2775-9237

DOI: 10.31014/aior.1992.05.03.437

Predicting Household Resilience Before and During Pandemic with Classifier Algorithms*

Ndari Surjaningsih[1], Hesti Werdaningtyas[1], Faizal Rahman[1], Romadhon Falaqh[2]

[1] Macroprudential Policy Department, Central Bank of Indonesia, Jakarta, Indonesia
[2] Research Assistant at Macroprudential Policy Department, Central Bank of Indonesia, Jakarta, Indonesia

Correspondence: Ndari Surjaningsih, Macroprudential Policy Department, Central Bank of Indonesia, Jakarta, 10110, Indonesia. E-mail: ndari@bi.go.id

Abstract
One of the lessons learned from the global financial crisis in 2008 was raising attention to monitoring and maintaining household vulnerability, particularly household credit risk, by using the default rate as the indicator. The indicator would be worsening at the economic recession, likewise, recently happened caused by the pandemic. The default event has a complex nonlinearity relationship among the determinants. To tackle the complex relationship, this study suggests exploiting machine learning approach in modeling the probability of default, especially the individual and ensemble classifiers. Therefore, this study aims to investigate changes of the Indonesian household financial resilience before and during the pandemic, supported by the individual-level data of the Financial Information Service System. This study finds that the ensemble classifiers, notably extreme gradient boosting, have a more predominant performance than the individual classifiers. The best model, then has the feature importance analysis to identify the variable pattern in explaining the default event periodically which reveals the pattern changes before and during the pandemic. The cost of debt/repayment capability and the policy mix is significant in explaining the default event. At the same time, the project location feature weakens in discriminating the target class.

Keywords: Default Event, Household Resilience, Vulnerability, Machine Learning

1. Introduction

The global financial crisis in 2008 has opened the authorities' attention to closely monitor household sector's vulnerability originated from uncontrolled household credit, particularly the housing loans. Among other indicators, the non-performing loans (NPL) could be a sign of vulnerability in term of household credit risk. In order to anticipate further deterioration of the NPL, it is then important to study factors determine the NPL or default rate. The study by Ali and Daly (2010) concluded that the contracted economy significantly increases the NPL rate. In line with this study, COVID-19 pandemic led to economic recession which would exacerbate household credit risk and vulnerability. Despite macroeconomic factor, household characteristics could be as the determinant of NPL, for instance, debt to asset, debt to income, occupation, age, and gender (Albacete and Linder,

* The results and opinions expressed in this paper are the authors' own and do not necessarily represent those of the Central Bank of Indonesia

2013; Deng and Liu, 2008; Shi et al., 2013; Alvaro and Gallardo, 2012; Feng et al., 2018; Jadhav et al., 2018; Barbaglia et al., 2020).

As a mean to understand the complex nonlinearity relationship among the variables, Heo et al. (2020) and other studies (Tsai and Chen, 2010; Crook et al., 2007; Marques et al., 2013) suggested the use of machine learning approach. The approach provides a new technique for modelling relationship instead of conventional statistical method. The model generated by a machine learning algorithm, especially an ensemble algorithm, has higher accuracy and more stable predictive performance (Li and Sun, 2013). The studies for investigating individual credit risk with a machine learning approach recently focused on the credit scoring/probability of default (PoD) in P2P lending, i.e., Teply and Polena (2020) and Song et al., (2020). Yeh and Lien (2009) estimated the PoD of credit cards in Taiwan. Mainly, these studies always compared the algorithms/classifiers categorized as linear, nonlinear, and rule/tree-based algorithms to choose the best model.

Following the aforementioned studies, this research investigates household financial resilience, specifically the PoD indicator, using an individual-level dataset in Indonesia namely the Financial Information Service System or SLIK (*Sistem Layanan Informasi Keuangan*). Since it has the biggest portion and the highest NPL rate in the structure of household loans, this study decides to analyze the mortgage loan. This research will compare models constructed by selected classifiers, i.e., logistic regression, support vector machine, random forest, and extreme gradient boosting (XGB), with two model evaluations, namely specificity and sensitivity. Higher specificity, without losing too much sensitivity score, is the best outcome of the model. This study obtains the XGB as the best model due to its performance in moderating both higher specificity and sensitivity scores. Moreover, it finds any substantial changes in the pattern of variables in explaining the households' financial vulnerability before and during a pandemic.

2. Method

2.1 Data Preparation

This study utilizes individual debtor reports on mortgage debt of SLIK dataset. The dataset consists of individual debtor data on mortgage debt as of December 2019 until June 2021. For this study objective, the dataset is divided into three periods, i.e., the pre-pandemic period (December 2019) with the low NPL rate, the peak of the pandemic period (June 2020) with the higher NPL rate, and the recovery period (June 2021) with the lower NPL rate than its peak in June 2020. For simplicity, those three periods are denoted as dataset 1, 2, and 3, respectively.

With the intention of identifying default event, debtors are categorized based on their NPL status, which determined by their collectability score between 1 to 5. For debtors whose collectability score between 3 to 5 are categorized as debtor in default event (symbolized as 1), otherwise will be denoted as 0. In other words, we construct the dependent variable of NPL status as binary. Furthermore, as independent variable, we set feature variables classified into three groups, namely (1) each debtors' characteristic variation, (2) mortgage facilities vary by facilities of each mortgage debtor, and (3) the time-related conditions of debtors for each period. For more complete explanation for each group's features are shown in the Table 1 below.

Table 1: Three groups of the features

	Debtor characteristics	Mortgage facilities	Time-related conditions
1	group of age	property type	current installment value
2	gross income	contract-type	frequency of restructuring
3	field of occupation	credit facility order	remaining maturity
4		project location	
5		maturity	
6		interest rate	
7		plafond value	
8		initial installment value	
9		LTV ratio	

2.2 The Selected Classifiers

This study applies the classification algorithm/classifiers since the target variable is categorical. Four selected classifiers can be considered as individual and ensemble identifiers or, respectively, as non-tree and tree-based algorithms. The first identifiers are logistic regression (LR) and support vector machine (SVM), then the last identifiers are random forest (RF) and extreme gradient boosting/xgboost (XGB). The four have a distinct set of hyperparameters that construct their model and provide the best testing score without overfitting in the training model.

2.2.1 Individual Classifiers

The individual identifiers, such as LR and SVM, could perform without aggregating/iterating sub-models to generate a model. To construct the model, LR determines the sigmoid function, while SVM sets the decision boundary, so that the model can classify the class of target variable (Yildirim, 2020). The studies commonly used both as the comparison model on credit scoring, such as Baesens et al. (2003), Bahnsen et al. (2014), Zekic-Susac et al. (2005), and Bensic et al. (2006). Nevertheless, both identifiers occasionally provide poorer performance against other algorithms i.e., neural networks and decision-tree based. Weaker performance might be caused by limited hyperparameter set of LR and SVM available for tuning, thus limiting the model to achieve the best outcome. Even this study only used three algorithm parameters for tuning the model.

2.2.2 Ensemble Classifiers

RF and XGB are considered ensemble identifiers since the models are generated using multiple identifiers in a particular technique. There are three techniques for aggregating/iterating multiple identifiers: bagging (bootstrap aggregating), boosting, and stacking (Massaoudi, 2020). For instance, the RF algorithm uses bagging to construct one model after training each sub-identifier with a different subset of training data. Meanwhile, XGB generates model by boosting as iterative method. In other words, boosting objective is to train in a corrective way one identifier that depends on poorer performance of another identifier from the previous iteration. Each identifier uses a complete set of training data. The stacking technique aggregates the result obtained from different algorithms, such as LR, XGB, and RF, into a conclusive outcome.

Some studies employ the ensemble identifiers in their comparison of credit scoring models, i.e., Brown and Mues (2012), Trivedi (2020), and Chopra and Bhilare (2018). They found that the ensemble learners outperform the individual ones with higher accuracy. Better performance of the ensemble learners might be due to availability of more hyperparameters for tuning the model outcome. The study used six and seven parameters of RF and XGB, respectively, in the tuning process.

2.3 Model Evaluation

To determine the best model, this study needs the appropriate model evaluation. The NPL dataset encounters imbalanced dataset issue caused by the number of default debtors, whose NPL is categorized as 1 or positive class, is too small than another class (categorized as 0/negative class). Considering such dataset limitation, this study does not apply the accuracy criteria for model evaluation. For this purpose, thus we employ sensitivity and specificity due to their capability to score model performance in balancing the outcome of true positive and true negative, which occasionally becomes a trade-off, particularly in the imbalanced target classes. The aim of model evaluation is to attain high sensitivity and specificity.

2.3.1 Sensitivity

Sensitivity or recall is a score that shows how sensitive the model is in identifying the true positive. In other words, sensitivity will assist this study in better identifying the model performance by minimizing type II error or false negative. The scoring formula is:

$$sensitivity = \frac{true\ positive}{(false\ negative + true\ positive)} \quad (1)$$

2.3.2 Specificity

Specificity is a score indicating the model's ability to identify the true negative. Moreover, this score can reveal the model performance to minimize the type I error or the false positive. The scoring formula is:

$$spesificity = \frac{true\ negative}{(false\ positive + true\ negative)} \quad (2)$$

3. Results

3.1 Best Modeling

This study splits the dataset into training and testing data subsets. The training data is used for constructing the model with a particular algorithm, namely LR, SVM, RF, and XGB. Then, this study evaluates the model produced by these algorithms with respect to both training and testing data subsets. Therefore, each algorithm consists of three models, which refers to the three datasets being used. The best outcome of each algorithm/model is chosen based on the discrepancies between training and testing scores. Low discrepancy indicates the model is not overfitting in the training data.

Table 2: Model Evaluation

Score: Sensitivity (Specificity)	Dataset 1 (Dec 2019)		Dataset 2 (June 2020)		Dataset 3 (June 2021)	
Algorithm/Model	Train	Test	Train	Test	Train	Test
LR	0.64 (0.71)	0.64 (0.71)	0.62 (0.65)	0.62 (0.65)	0.64 (0.71)	0.64 (0.71)
SVM	0.70 (0.24)	0.69 (0.24)	0.85 (0.13)	0.85 (0.13)	0.95 (0.03)	0.96 (0.03)
RF	0.70 (0.70)	0.70 (0.70)	0.63 (0.70)	0.62 (0.70)	0.75 (0.75)	0.75 (0.75)
XGB	0.71 (0.72)	0.71 (0.72)	0.70 (0.69)	0.69 (0.69)	0.76 (0.77)	0.76 (0.77)

According to Table 2, all models reach their best outcome, as shown by small discrepancy between the training and testing scores. Nevertheless, this study selects the best algorithm that provides the highest sensitivity score along with not too low specificity score. The XGB algorithm is considered as the best model because of its predominant performance in all datasets. Even though they have the highest sensitivity score in datasets 2 and 3, the SVM models fail to achieve the same high score in specificity, or in other words they do not fulfil the aim of this model evaluation. The inferior performance of SVM implies that XGB has a capability to balance the two scores. Another ensemble classifier, the RF algorithm, could also moderate the two scores, however, its performance cannot compete with the high sensitivity score of XGB. Instead of the individual classifiers, the ensemble classifiers are proper as algorithms to model the imbalance classes, particularly of default/NPL in this study.

3.2 PoD Determinants from Dataset 1 (Pre-pandemic Period)

In analyzing the determinants of PoD, we use the feature importance (FI) analysis. The FI is a method to identify features that have significant factors in affecting model performance if those features are experimentally randomized. Subsequently, the result of this analysis is namely the mean decrease accuracy (MDA). The higher MDA, the more influential the feature is for the model. Table 3 shows all features sorted by the MDA of XGB models of Dataset 1 and others.

Table 3: Feature importance

	Dataset 1	Dataset 2	Dataset 3
1	project location	interest rate	interest rate
2	remaining maturity	current installment value	current installment value
3	maturity	project location	frequency of restructuring
4	frequency of restructuring	maturity	remaining maturity
5	interest rate	remaining maturity	maturity
6	current installment value	initial installment value	LTV ratio
7	initial installment value	frequency of restructuring	project location
8	group of age	gross income	group of age
9	plafond value	plafond value	field of occupation
10	gross income	group of age	initial installment value
11	field of occupation	contract-type	contract-type
12		field of occupation	gross income
13		credit facility order	plafond value
14			credit facility order
15			property type

Four features in the XGB model of Dataset 1 are dropped to maintain the best outcome of model performance. The excluded features are property type, LTV ratio, credit facility order, and contract type. As presented in Table 3, the most important feature for the model of Dataset 1/pre-pandemic period is project location. Six regions are included in the project location feature, namely Java, Sumatra, Bali-Nusa Tenggara Islands (Balinusra), Borneo, Sulawesi-Moluccas-Papua (Sulampua), and Others. The most important feature implies that the PoD of mortgage debtors tends to be high in a specific region during the pre-pandemic period. Based on project location, the NPL rate outside Java records an average of 4.77%, which is higher than in Java (1.87%). Moreover, the rest of the top three features are the remaining maturity and the maturity, which indicate that the shorter the remaining maturity and the longer maturity of the mortgage, the higher the PoD of the debtors. The last three features in the top six key features are frequency of restructuring, interest rate, and current installment value, denoting the higher of those features, the higher the PoD of the mortgage debtors.

3.3 Analysis of PoD Determinants Pattern Changes due to the pandemic

Based on the FI in Table 3, the rank or position of interest rate during the pandemic escalates to the most important feature in both periods (datasets 2 and 3). Moreover, the current installment value feature also becomes more important feature during the pandemic period. It goes up to the second-best in dataset 2, as well as in the third dataset or period. The first position feature in pre-pandemic period (datasets 1), namely project location, weakens to the third rank in the peak of pandemic period (datasets 2) and suffers further decline to the seventh rank in the recovery period (datasets 3). The attenuation of the project location feature is in the same direction of narrowing discrepancy of the NPL rate between Java and outside Java after the Covid-19 outbreak struck all regions entirely. That narrowing of discrepancy lessens the ability of project location to accurately classify the default event, given the widespread outbreak successfully increases the PoD of debtors regardless the mortgage project location.

For the best outcome, the XGB model of dataset 2 omits two features, namely LTV ratio and property type. The omitted features, nonetheless, do not belong to the model of dataset 3. Even the LTV ratio, which previously does not account for the first two models, surprisingly turns out to become the sixth-best feature, and the frequency of restructuring jumps to the third position in the model of dataset 3. The change of variables/features indicate that there are two factors significantly impact debtors' collectability during a pandemic. These factors are the cost of debts (repayment capability) and the policy mix. The interest rate which applies as the cost of debts and current installment value, consistently culminate as the top two of feature importance during the pandemic periods. It shows that both features are increasingly significant to the debtors' repayment capability, thus accordingly, the debtors' collectability is more sensitive to both features. Furthermore, the economic recession in the early pandemic era, such as the negative income shock for the debtors, has strengthened the impact exceedingly (Ali and Daly, 2010). On the other hand, the rising importance of both features, i.e., the frequency of restructuring and the LTV ratio, represent that the policy mix, including the loans restructuring and the loosening macroprudential policy through the LTV instrument, could restrain the escalating PoD of mortgage debtors in Indonesia continually, especially at the period of recovery after the peak of the pandemic impact in June 2020.

4. Conclusion

The study suggests that the ensemble classifiers, notably the XGB algorithm, are appropriate for modelling the imbalance target class, especially on the NPL/default event. The ensemble identifiers have superior capability to achieve balanced high scores, both in sensitivity and specificity than the individual ones. Trade-off between these two scores is occasionally happened in the SVM algorithm, in the sense that the SVM may produce higher sensitivity but with an extremely low specificity. The best model can reveal the feature importance for each period to understand more precisely how the features explain the PoD. In the first/pre-pandemic period, the feature importance analysis tells us that the project location becomes the most important feature for identifying factors behind the default event. There had been a wide discrepancy in terms of default rate between Java Island and outside Java Island until the outbreak struck all regions. Project location feature falls off to the third position in June 2020 and degrades to seventh rank in the next period. The interest rate afterwards succeeds as the first position, and the current installment feature improves to the second place of the feature importance at the last two consecutively periods. Moreover, the frequency of restructuring and the LTV ratio are also gaining better positions in the feature importance analysis, particularly during the recovery period. The dynamic in the variables or feature importance articulates that the cost of debts (implying the debtors' repayment capability) and the benefits of the policy mix substantially impact the quality of collectability or PoD of mortgage debtors in the pandemic era.

References

Albacete, N., & Lindner, P. (2013). Household Vulnerability in Austria–A Microeconomic Analysis Based on the Household Finance and Consumption Survey. *Financial Stability Report*, 57-73.

Alfaro, R., & Gallardo, N. (2012). The determinants of household debt default.

Baesens, B., Van Gestel, T., Viaene, S., Stepanova, M., Suykens, J., & Vanthienen, J. (2003). Benchmarking state-of-the-art classification algorithms for credit scoring. *Journal of the operational research society*, *54*(6), 627-635. DOI: 10.1057/palgrave.JORS .2601545

Bahnsen, A. C., Aouada, D., & Ottersten, B. (2014, December). Example-dependent cost-sensitive logistic regression for credit scoring. In *2014 13th International conference on machine learning and applications* (pp. 263-269). IEEE.

Barbaglia, L., Manzan, S., & Tosetti, E. (2020). Forecasting loan default in Europe with machine learning. *Available at SSRN 3605449*.

Bensic, M., Sarlija, N., & Zekic-Susac, M. (2005). Modelling small-business credit scoring using logistic regression, neural networks, and decision trees. *Intelligent Systems in Accounting, Finance & Management: International Journal*, *13*(3), 133-150.

Brown, I., & Mues, C. (2012). An experimental comparison of classification algorithms for imbalanced credit scoring data sets. *Expert Systems with Applications*, *39*(3), 3446-3453.

Chopra, A., & Bhilare, P. (2018). Application of ensemble models in credit scoring models. *Business Perspectives and Research*, *6*(2), 129-141.

Crook, J., Edelman, D., & Thomas, L. (2007). Recent developments in consumer credit risk assessment. *Eur. J. Oper. Res. 183*, 1447–1465

Deng, Y., & Liu, P. (2009). In China's housing market, mortgage prepayment and default behaviour with embedded forward contract risks. *The Journal of Real Estate Finance and Economics, 38(3)*, 214-240.

Feng, X., Xiao, Z., Zhong, B., Qiu, J., & Dong, Y. (2018). *Appl. Soft Comput. (ISSN:1568-4946) 65*, 139–151.

Heo, W., Lee, J. M., Park, N., & Grable, J. E. (2020). Using Artificial Neural Network techniques improves the description and prediction of household financial ratios. *Journal of Behavioral and Experimental Finance, 25, 100273*.

Jadhav, S., He, H., & Jenkins, K. (2018). Information gain directed genetic algorithm wrapper feature selection for credit rating. *Appl. Soft Comput. (ISSN: 1568-4946)*.

Li, H., & Sun, J. (2013). Predicting business failure using an RSF-based case-based reasoning ensemble forecasting method. *J. Forecasting 32*, 180–192.

Marques, A., Garcia, V., & Sanchez, J. (2013). A literature review on the application of evolutionary computing to credit scoring. *J. Oper. Res. Soc. 64*, 1384–1399

Massaoudi, T. (2020). *What You Should Know About Ensemble Learning*. Retrieved from Towards Data Science: https://towardsdatascience.com/what-you-should-know-about-ensemble-learning-e92d4b3c3608

Shi, J., Zhang, S. Y., & Qiu, L. M. (2013). Credit scoring by feature-weighted support vector machines. *J. Zhejiang Univ. Sci. C 14 (3)*, 197–204.

Teply, P., & Polena, M. (2020). Best classification algorithms in peer-to-peer lending. *The North American Journal of Economics and Finance*, *51*. DOI: 10.1016/j.najef.2019.01.001

Trivedi, S. K. (2020). A study on credit scoring modelling with different feature selection and machine learning approaches. *Technology in Society*, *63*, 101413.

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

Yeh, I. C., & Lien, C. H. (2009). The comparisons of data mining techniques for the predictive accuracy of probability of default of credit card clients. *Expert Systems with Applications, 36(2)*, 2473-2480.

Yildrim, S. (2020). *11 Most Common Machine Learning Algorithms Explained in a Nutshell*. Retrieved from Towards Data Science: https://towardsdatascience.com/11-most-common-machine-learning-algorithms-explained-in-a-nutshell-cc6e98df93be

Zekic-Susac, M., Sarlija, N., & Bensic, M. (2004, June). Small business credit scoring: a comparison of logistic regression, neural network, and decision tree models. In *26th International Conference on Information Technology Interfaces, 2004.* (pp. 265-270). IEEE.