ADABOOST-SVM AND FEATURE SELECTION OF GENETIC ALGORITHM COMBINATION TO ENHANCE INDONESIAN P2P LENDING CREDIT RISK ASSESSMENT

Devon Ho and Ira Fachira

School of Business and Management, Institut Teknologi Bandung, Indonesia
Email: devon.ho@sbm.itb.ac.id

Abstract. Many Fintech start-ups were established over the past few years in Indonesia. They saw the opportunity that SMEs needed them as platform to provide loans and fundings in order to run their businesses. However, in that sense, many Fintech companies were considered failed to payback its lenders because they could not gain a sufficient amount of borrowers to join the P2P lending platform. As a result, they experienced an increase in loan loss rate, as well as non-performing loans ratio. That was the reason why many Fintech could not survive and were forced to close their business. Well-performing credit risk management is one of the work that Fintech companies should do. Machine learnings have already been utilized to enhance credit risk assessment, although they still need improvement following the development and changes towards modernization. Some of the techniques such as AdaBoost-SVM and Genetic Algorithm will be discussed in this paper and the author will analyse the effectiveness of both techniques and discover the possibility of having them combined to generate the best outcome.

Keywords: Fintech; non-performing loans ratio; *AdaBoost-SVM; Genetic Algorithm;credit risk assessment; machine learnings.*

INTRODUCTION

Since the development of Fintech in Indonesia was exponentially increased in the past few years, many Fintech peer-to-peer lending companies were experienced some difficulties, for example, when the loan deadline has expired, customers often just disappear. As a platform that is trusted by borrowers, Fintech companies should collect the debts from their customers. Only, there are times when the borrower is in a long distance, then the borrower would give up the money that has been disbursed.one of them was to determine the suitable credit risk assessment. Inability to maintain the non-performing loans ratio, experiencing huge losses, and failed to satisfy its customers are the main drivers that forced some of the Fintech companies to shut down. It is mandatory that credit risk assessment should be performed well by Fintech companies, in order to have a sustainable businesses. When such companies have assessed the credit risk well, they can create healthy relationship with its stakeholders, and that include their customers. One of the solution proposed to react to this problem is to have certain guarantor that the credit risk can be minimized, and machine learning is the one that have been discovered and developed as a virtual helper to analyse the best method that can be applied when dealing such situation.According to Moody's Analytics(2017), a machine learning model, unconstrained by some assumptions pf classical statistical models, can yield much better insights that a human analyst could not infer from the data. Machine learning has been known to have different nature and reaction towards the problem, and it is important to acknowledge some of them. This paper will introduce how the machine learning methods such as Support Vector Machine and Genetic Algorithm can be utilized, as well as analysing the combination approach between two machine learning would create better assessment towards credit risk.

LITERATURE REVIEW

Machine learning is a way of computers to process data, learn from it, and then make a decision or prediction regarding new data. To accomplish a task, the machine contains itself with a large amount of data and algorithm, and learn how to perform the task. Statistical learning is one of the ability of the machine learning that is carried within the program. The goal is to find and learn from patterns and trends within large datasets to predict upcoming occurrence. The development of machine learning has been through a complex phase. According to James AI(2017), There are two main types of machine learning: the first type of learning is called supervised learning. Supervised learning allows the input data to describe the characteristics of given individuals, which later on will be determined by labelling to the output of the data. The algorithm would study from the association between the label and the characteristics and is able to generate prediction of the label of an individual given the characteristics. The second type of learning is called unsupervised learning. This type of machine learning requires no labels to the algorithm who has to find structure in its input on its own.As for credit risk, focusing on supervised learning are the best option since the loans are labelled (in this case, "default/safe" or "good/bad") and it is fit with the goal of the paper to predict whether borrower is likely to default or not, given their characteristics. Being recognized as one of the first domains to drive interest in using artificial intelligence, even before high computing machines were available. A lot of vintage mathematic

formulas were being utilized by most of the researchers at that time until now, to construct the brilliant solution in forms of AIs, such as machine learning, deep learning, data mining(Medium,Sydney,2018). Many Fintech companies in several countries have applied AI to their system, in it worked better than ever before.

When such customer in this case, Small and Medium Enterprises, wants to apply for a loan to lending platforms, this lending platform or Fintech companies must be able to evaluate whether the applicants have the ability to repay the loan at the right time by looking at their portfolio and track record, so that the default risk could be minimized. Profitability ratio calculation and leveraging are used by lending platforms to assess credit risk.A profitable firm are expected to have good liquidity, meaning that they can generate enough cash to cover interest expense and principal due. However, firm that are being leveraged more, has insufficient equity to cover sudden economic shocks. Given this scenario, there are two loan applicants, first one is having high profitability and high leverage, while the other has a low profitability and low leverage, which firm has lower credit risk? It is a quite complicated question to be answered of course. It is even complicated when such financial institutions gather the other dimensions to be examined alongside with the portfolio when performing credit risk assessment. These additional dimensions appear to be other financial information such as solvency ratio, or behavioral information such as loan/trade credit payment behavior. If traditional statistical learning is used, it will require more time and definitely be much challenging, so machine learning really helps to provide easiness to companies in terms of achieving goals(Bacham D,2017).

In Indonesia, AI should be considered as a more effective and efficient methods to overcome the problems, given that Indonesia is filled with vast population, and Fintech companies need something that can cover up those number with just short amount of time. It is right time for them to try developing the AI as brand new solution by trying to apply them out with provided methods. Many researcher are trying to figure out how to apply this existing algorithm methods in proper with credit risk assessment, so that Fintech companies can perform this assessment efficiently.

METHODOLOGY

Desk research will be used as research method of this paper. Sources are collected from academic journals, theories, previous research, and articles that are related to the research and will be cited as references. The author would first present the background of the problem, in this case, the difficulties of Fintech P2P lending platform to assess credit risk, which rises the next question, which tool that is ideal to help Fintech platforms assessing their credit risk.Then, it turned out that machine learning (AdaBoost-Support Vector Machine and Genetic Algorithm) is one of the best solution to overcome the problem. By introducing them at a glimpse, readers would at least have an understanding on how magnificent these machine learnings to be studied. The key point of the research is that, the author would generate comprehensive findings on what if such machine learnings techniques are combined, and how would the result represent the new solution for Indonesia P2P lending platform. Through the findings, conclusion and recommendation can be made for further research.

FINDINGS AND ARGUMENT

Below are some of the machine learning methods and features the capability of generating better accuracy in terms of credit risk assessment:

1. AdaBoost-Support Vector Machine method

Being recognized as one of the best machine learning technique because of the high accuracy,AdaBoost has become popular to be developed by scientist. In 2005, Caruana & al. made an empirical comparison of supervised learning algorithms. They included Random Forest and Boosted decision trees and concluded that boosted trees were the best learning algorithm overall, being random forest are closed to second place. Researchers from Dian Nuswantoro University in Indonesia, named Defri Kurniawan and Catur Supriyanto developed the idea to increase the accuracy of machine learning algorithm which is SVM(Support Vector Machine) by applying ensemble method that consist of two elements, boosting and bagging. Bagging and Boosting have the same approach in making decisions with using multiple sounds combined into a single prediction. But the difference lies from the way they took that single prediction. In Bagging, the model receives the same weight from the most votes. Whereas Boosting, weights are used to give effect to further weight improvements. Here, the research was utilizing dataset from Australia Credit Approval as a subject of the research. Closeness and similarity of the dataset are the reasons why they choosed Australian dataset.

Figure 1. Results generated from different SVM method

From the dataset, different results were generated by also performing different kind of methods. Above shows the result of how AdaBoost-SVM method can increase accuracy of the credit risk assessment, with the number reaching 86.09 percent.

2.	Feature Selection of Genetic Algorithm-Neural Network.

Feature selection is a pre-processing technique that is commonly used on high-dimensional data, and its purposes include reducing the dimensionality, removing irrelevant and redundant features, facilitating data understanding, reducing the amount of data needed for learning, improving the predictive accuracy of the algorithms, and increasing the interpretability of the models (Oreski, 2012). Feature selection is the problem of choosing a small subset of features that ideally is necessary and sufficient for describing the target concept (Kira & Rendell,1992). Therefore, it is important to perform feature selection well in order to have a well-set data, as well as eliminating the irrelevant information in order to prevent inaccuracy while processing the datasets. Researchers named Stepan and Goran Oreski decided to propose this sophisticated feature selection to increase effectiveness and efficiency in data processing. Genetic algorithm was used as a tool for enhancement in feature selection and classification in credit risk assessment process. Here, they used hybrid combination between Genetic Algorithm and Neural Network, which is often called HGA-NN. Assume that Indonesian credit dataset is the same with German and Australian dataset, below are the result of how these machine learning method generate different results, in this case, accuracy of credit risk assessment.
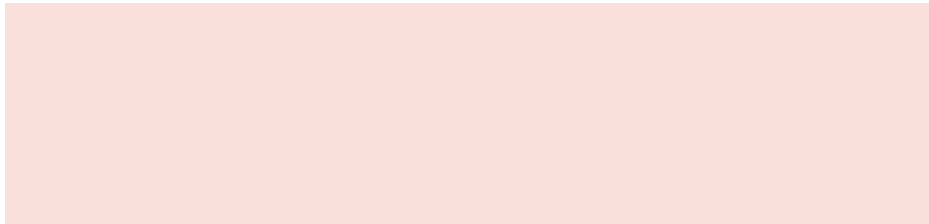
CONCLUSIONS

In conclusion, machine learning provides a better outcome to become solutions towards credit risk assessment. All of the machine learning method generate good results in term of accuracy. AdaBoost has been recognised as being one of the best among the other methods and it is even better if it is combined all and all with the feature selection of Genetic Algorithm, which increase the accuracy to detect errors and possibility of default. Regardless, this findings has been shown to have limitation, such as research scope and sample. At least the findings had shown that the solution is exist. Therefore, it can be said that Indonesian P2P lending could use this as a tool for better credit risk assessment in the near future.

REFERENCES

Bacham D., Zhao J.(2017). *Machine Learning: Challenges,Lessons, and Opportunity in Credit Risk Modelling.* Available at https://www.moodysanalytics.com/risk-perspectivesmagazine/managing-disruption/spotlight/machine-learning-challenges-lessons-and-opportunitiesin-credit-risk-modeling.

Bharadwaj R.(2019).*Machine Learning for Credit Risk- What's Changing, and What Does It Mean?*.Available at https://emerj.com/ai-podcast-interviews/machine-learning-for-credit-risk/.

CNN Indonesia.OJK Catat Rasio Kredit Macet Fintech per Februari 3,18%. Available at https://www.cnnindonesia.com/ekonomi/20190328200018-78-381617/ojk-catat-rasio-kredit-macet-fintech-per-februari-318-persen

James(2017).*Machine Learning in Credit Risk Modelling:Efficiency should not come at the expense of Explainability*.

Kurniawan D., Supriyanto C.(2013).*Optimasi Algoritma Support Vector Machine(SVM) Menggunakan Metode Adaboost Untuk Penilaian Risiko Kredit.* Jurnal Teknologi Informasi Volume 9 Nomor 1.

Oreski S.(2014).*Genetic algorithm-based heuristic for feature selection in credit risk assessment*. Expert System with Applications 41(2014), 2052-2064.