

A Review of Contribution and Challenge in Predictive Machine Learning Model at Financial Industry

Irman Hermadi¹, Yani Nurhadryani¹, Indra Ranggadara^{1*}, and Ruhul Amin¹

¹Department of Computer Science, Faculty of Mathematics and Natural Sciences, Bogor Agricultural University, 16680 Bogor, Indonesia

*indra.ranggadara@mercubuana.ac.id

Abstract. Prediction is essential one of them in the financial industry, with the prediction of this industry to be confident in determining the next steps to develop business. In the era of developing technology now, the application of machine learning to help the industry is essential, not only helps to count alone but can also help industry players to be able to predict in cases faced in the industry. In Machine Learning, financial industry players can create a credit risk analysis model based on available alternative data so that they can better manage overall risk. In this study aims to provide information on the contribution and challenges of machine learning in the financial industry and generate information on how the contributions and challenges in the application of machine learning have compiled from other researchers.

1. Background

Innovations in information technology and their supporters have developed very rapidly and will continue to develop over time. So, almost every aspect and field of science cannot separated from IT. in the field of finance or the financial industry. Innovation in technology has a significant role in the financial industry because most of the business processes and entirely managed by computer systems. The use of technology in the financial sector or the financial industry is also proliferating from administration to overseeing the flow of transactions to reduce transaction fraud and other illegal transaction activities. In this case, the financial industry has shifted to machine learning to improve performance in every business process. Machine learning can broadly defined as a computational method that is used based on experience. The term machine learning is a computer process for learning from data. Without data, the computer will not be able to learn anything. Therefore if we want to learn machine learning, we will continue to interact with data. All machine learning knowledge will involve data. Data can be the same, but the algorithm and the approach are different to get optimal results. In this case, the experience referred to is previous information in the form of electronic data that was collected and provided as material for analysis. The data is digital data that has been labelled by humans themselves or data obtained from interactions with the environment. Then the method of using machine learning for the financial industry also varies depending on the purpose of the industry itself, such as Support Vector Machine, or even combining machine learning techniques with other models such as using machine learning analysis to predict interest curves. However, the majority is used to make



accurate predictions. For the contribution in this research only focuses on summarization conducted by researchers.

2. Machine Learning

Machine Learning is a study of algorithms to learn something in doing certain things done by humans automatically[1]. Machine learning techniques generally focus on finding connections in the observed data and mining those relationships[2]. Machine Learning is one of the fields included in artificial intelligence that can affect various other aspects, namely static, mathematics, and various theoretical aspects of computer science. Machine Learning has the goal to learn an algorithm to do a learning system automatically with a very minimal contribution made by humans in general. Learning in this case relates to completing various tasks that exist, or making a prediction of accurate new conclusions from various patterns[3] that have been studied previously. Here are some examples of algorithms from the concept of Supervised Learning and Unsupervised Learning, according to Smola [4]:

- a. Supervised Learning; is learning based on a set of examples of the desired input and output pairs[5]. This algorithm observes these examples and then produces a model that can map the new input into the right output while.
- b. Unsupervised Learning; aims to learn and look for unusual patterns in the input provided[3], Although not explicitly provided the right output. One of the most commonly used unsupervised learning algorithms is clustering.
- c. Semi-Supervised Learning; It combines supervised with unsupervised to produce a function. Russel [5] explain that the semi-directed learning algorithm given examples of the right number of inputs and outputs and a set of inputs whose outputs are not yet known. This algorithm must create a series of units between two types of algorithms to be able to cover weaknesses in each algorithm.

3. Research Review

This review will discuss at least 5 studies related to machine learning in financial industry. Broadly speaking will discuss the contribution and challenge of the research conducted. In this discussion does not discuss specific factors that are outside the topic of machine learning, but will only summarize the research.

4. Result and Discussion

This section explains the results of a brief review of the implementation and obstacle in 5 studies related to machine learning. Here, objectively summarizing the contribution and the challenge faced, and the research that reviewed randomly can see in Table 1.

Table 1. Review result

| Authors | Contribution | Challenge |
|---|---|---|
| Yachao Li and Yufa Wang [6] | Compare the statistical method and machine learning method to predict bankruptcy with utilizing China listed companies | Different indicators may have different characteristics and not all indicator scan be analysed |
| Ong Shu Yee, Saravanan Sagadevan and Nurul Hashimah Ahamed HassainMalim [7] | Machine learning (ML) techniques were employed to predict the suspicious and non-suspicious transactions automatically by using classifiers | Tested classification metrics by using five Bayesian classifiers namely Naïve Bayes, K2, TAN, Logistics and J48 |

| Authors | Contribution | Challenge |
|---|--|--|
| Kou, Gang, Chao, Xiangrui, Peng, YiAlsaadi, Fawaz E.Herrera-Viedma, Enrique [8] | To introduce current researches on financial systemic risk with machine learning methods | Lots of traditional econometrics and statistical models are also used to systemic risk analysis |
| Mikhail Kanevski and Vadim Timonin [9] | Review of the analysis and modelling of Swiss franc interest rate curves (IRC) by using unsupervised (SOM, Gaussian Mixtures) and supervised machine (MLP) learning algorithms | Analysis of local cross-correlations along with clustering in model's parameter space and mapping of interest rate for the visualization, missing data recovering and forecasting purposes |
| Lkhagvadorj Munkhdalai, Tsendsuren Munkhdalai, Oyun-Erdene Namsrai, Jong Yun Lee and Keun Ho Ryu [10] | an automated credit decision-making system that can avoid loss of opportunity or credit losses to reduce potential loss for each lending institution. | the prior work focused only on their performance in binary credit classification. It is inefficient and not practical from the perspective of the banking risk management |

For the first research from Yachao Li and Yufa Wang's [6], explain they compare the statistical method and machine learning method to predict bankruptcy with utilizing China listed companies, because Much bankruptcy prediction has relied on statistical models and got low prediction accuracy. Second research from Ong Shu Yee, Saravanan Sagadevan and Nurul Hashimah Ahamed Hassain Mali's [7], the patterns and characteristics of suspicious and non-suspicious transactions based on normalized and anomalies data, used the supervised based classification using Bayesian network classifiers namely K2, Tree Augmented Naïve Bayes (TAN), and Naïve Bayes, logistics and J48 classifiers. Third Research from Kou, Gang, Chao, Xiangrui, Peng, YiAlsaadi, Fawaz E.Herrera-Viedma, Enrique's [8], explain to detect and respond to systemic risk with growing amounts of data produced in financial markets and systems, and they survey existing researches and methodologies on assessment and measurement of financial systemic risk combined with machine learning technologies, including big data analysis, network analysis and sentiment analysis. Fourth research from Mikhail Kanevski and Vadim Timonin's [9], to deals with the review of the analysis and modeling of Swiss franc interest rate curves (IRC) by using unsupervised (SOM, Gaussian Mixtures) and supervised machine (MLP) learning algorithms, analysis of NSM parameters and their temporal and clustering structures helps to understand the relevance of model and its potential use for the forecasting. Then the last from Lkhagvadorj Munkhdalai, Tsendsuren Munkhdalai, Oyun-Erdene Namsrai, Jong Yun Lee and Keun Ho Ryu's [10] explain to establish a new benchmark using real consumer data and to provide machine-learning approaches that can serve as a baseline on this benchmark. We performed an extensive comparison between the machine-learning approaches and a human expert-based model—FICO credit scoring system—by using a Survey of Consumer Finances (SCF) data.

5. Summary

From several journals show that the implementation of machine learning in the financial industry, there are still many challenge, even though the use of machine learning also has a positive influence that can support business processes in the financial industry. Besides, machine learning in the financial industry

will continue to develop, along with technological advances and corporate strategies that must follow changes and circumstances. With the many challenges in implementing machine learning, it is hoped that there will be additional contributions by combining many techniques or applying new methods.

References

- [1] S. Shalev-Shwartz and S. Ben-David, *Understanding Machine Learning From Theory to Algorithm*. New York: Cambridge University Press, 2014.
- [2] B. M. Khammas, A. Monemi, J. Stephen Bassi, I. Ismail, S. Mohd Nor, and M. N. Marsono, "Feature Selection And Machine Learning Classification For Malware Detection," *Jurnal Teknologi*, vol. 77, no. 1, Oct. 2015.
- [3] K. P. Murphy, *Machine Learning A Probabilistic Perspective*. London: The MIT Press Cambridge, 2012.
- [4] V. Smola, *Introduction to Machine Learning*. New York: Cambridge University Press, 2008.
- [5] S. J. Russell, P. Norvig, J. F. Canny, J. M. Malik, and D. D. Edwards, *Artificial Intelligence A Modern Approach*. New Jersey: Prentice Hall, 1995.
- [6] Y. Li and Y. Wang, "Machine Learning Methods of Bankruptcy Prediction Using Accounting Ratios," *Open Journal of Business and Management*, vol. 06, no. 01, pp. 1–20, 2018.
- [7] O. S. Yee, S. Sagadevan, and N. H. A. H. Malim, "Credit Card Fraud Detection Using Machine Learning As Data Mining Technique," *Journal of Telecommunication, Electronic and Computer Engineering*, vol. 10, pp. 1–4, 2018.
- [8] G. Kou, X. Chao, Y. Peng, F. E. Alsaadi, and E. Herrera-Viedma, "Machine Learning Methods For Systemic Risk Analysis In Financial Sectors," *Technological and Economic Development of Economy*, vol. 25, no. 5, pp. 716–742, May 2019.
- [9] M. Kanevski and V. Timonin, "Machine learning analysis and modeling of interest rate curves," in *European Symposium on Artificial Neural Networks - Computational Intelligence and Machine Learning*, 2010.
- [10] L. Munkhdalai, T. Munkhdalai, O.-E. Namsrai, J. Lee, and K. Ryu, "An Empirical Comparison of Machine-Learning Methods on Bank Client Credit Assessments," *Sustainability*, vol. 11, no. 3, p. 699, Jan. 2019.