Explainability and Stability of Machine Learning Applications — A Financial Risk Management Perspective

Liyang Wang[1], Yu Cheng[2], Ningjing Sang[2], You Yao[3]

[1] Olin Business School, Washington University in St. Louis, St. Louis, MO, 63130, USA
[2] The Fu Foundation School of Engineering and Applied Science, Columbia University, New York, NY, 10027, USA
[3] Viterbi School of Engineering, University of Southern California, Los Angeles, California. 90089 USA
DOI: 10.32629/memf.v5i5.2902

Abstract: With advancement in computing power, hardware, and machine learning algorithms, more and more industry sectors have started to incorporate machine learning in the core business. The adoption of machine learning model in risk management is slower, due to the sensitive nature of the tasks, data involved, and regulatory pressure. This paper evaluates the explainability and stability of machine learning models on a traditional financial risk management task and found out that machine learning models can exhibit an enhanced level of adaptability and stability. However, different models could lead to drastically different performance, which require companies to spend additional resources in training and development. Overall, the net benefits are overwhelming, if done correctly.

***Keywords*:** machine learning, risk management, random forest, gradient boosting, bankruptcy

1. Introduction

This paper explores the application of machine learning from a financial risk management perspective. With rapid development in computing power and new algorithms, machine learning has become widely applicable. New breakthroughs such as large language model has shown the diversify of tasks AI can handle. Unlike many other sectors, financial risk management tends to adopt technology slower in favor of better explainability and higher stability. This paper shows that machine learning models can indeed achieve those but also do require careful tunning and possibly longer development cycle.

2. Literature Review

Machine learning model is one of the most notable technological breakthroughs in the recent decade and has revolutionized many aspects of the financial services industry. Even though initial adoption is slow, machine learning has become a vital tool and greatly improved operational efficiency [1].

In risk management, the adoption of machine learning model is slower compared to other lines of business, due to the risk-averse nature of the work [2]. Nonetheless, machine learning models have proven to be a great addition to the existing tools. For instance, previous studies have shown machine learning models are effective at identifying frauds more efficiently than conventional approaches [3]. Meanwhile, studies also argue that advanced machine learning algorithms such as neural networks can detect credit risks early and help banks in loss prevention [4]. Compared to conventional method, machine learning models allow banks to evaluate risks more holistically, not only due to numeric data [5]. Methods such as natural language processing enable risk assessment on qualitative data, which is traditionally undervalued and overlooked [6].

Overall, machine learning model has proved to be able to solve incumbent issues in different industries and offers new insights [7]. It has shown promises in sound analysis [8] and it is very promising to solve long-lasting issues in financial risk management.

Generally, machine learning is more prevalent in less regulated areas such as credit scoring and fraud detection, while the application is more restricted in highly regulated areas such as anti-money laundering. The major reason is due to stability and explainability concerns, which this paper aims to explore.

3. Research Methodology

This paper evaluates the performance stability and implications of the machine learning models in financial risk management. To achieve it, this paper fits several machine learning models to predict bankruptcy of publicly listed firms in the United States. The data is collected from publicly available sources. Firm-level financial information is collected from annual reports and stock prices data is from CRSP database. Some additional information like credit ratings are also collected from publicly available sources. The major machine learning model equation can be expressed below.

$$Bankruptcy_{i,t} = F_1\left(profit_{i,t}\right) + F_2\left(growth_{i,t}\right) + F_3\left(credit_{i,t}\right) + \ldots + F_n\left(sentiment_{i,t}\right)$$

The models to be fitted are logistic regression, random forest, and gradient boosting. Since bankruptcy is a highly imbalanced indicator, this paper also takes such characteristics into account and utilize down-sampling to enable smoother prediction.

An extended sensitivity analysis is conducted to evaluate if negatives identified by the machine learning models can be explained by firm performance. The extended bankruptcy label includes:

(1) Actual bankruptcy label

(2) Stock prices fall more than 90% in the last year and no signs of recovery

(3) Low liquidity and high leverage, along with high debt payables

(4) Negative equity/assets for a prolonged time period

The additional labels are firms that are near bankruptcy but has not filed for it officially. If machine learning models can cover those labels, it shows it can capture new risks, not identified by a set label. Thus, it should grant some confidence for applications for more sensitive topics.

4. Results and Implications

Different models have shown vastly different performances. Generally, tree-based models perform better than linear models. Figure 1 below shows some performance data for the models.

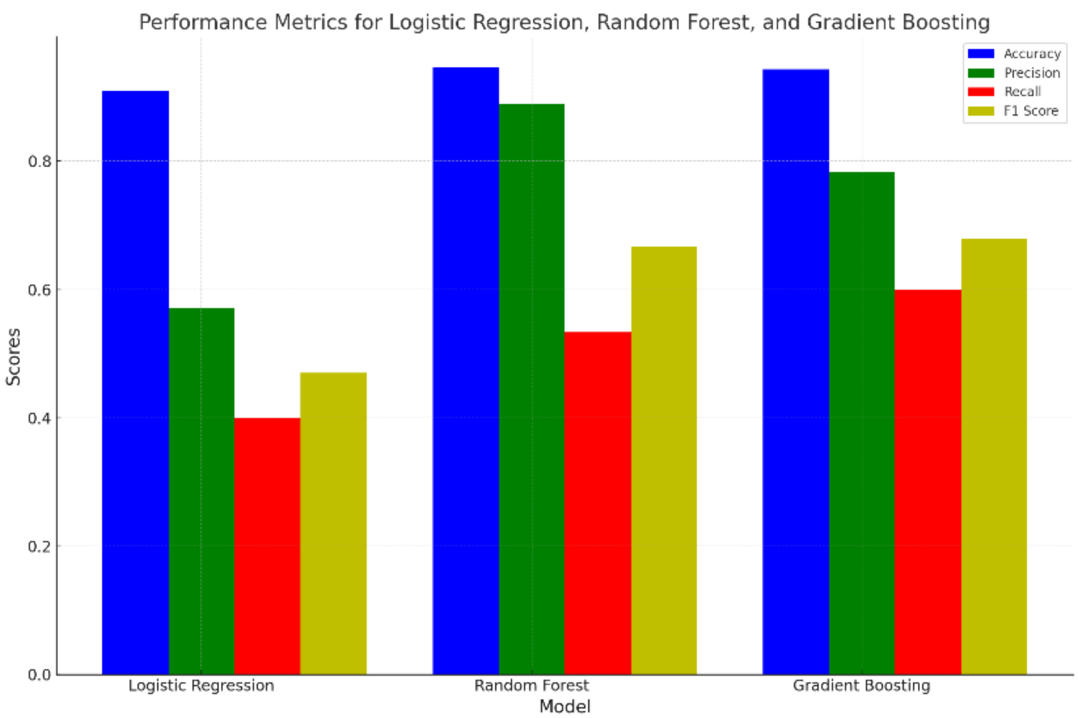

Figure 1. Performance Summary of the Models

All of the models have good prediction on accuracy. However, accuracy is not a good predictor for imbalanced dataset due to the imbalanced nature. Recall is the primary indicator. From the risk management perspective, a model needs to capture the truest positives, and random forest is the best performing model out of the three. Gradient boosting is the close second, even though it is slightly more efficient due to a higher precision, thus reducing the number of false positives in the process.

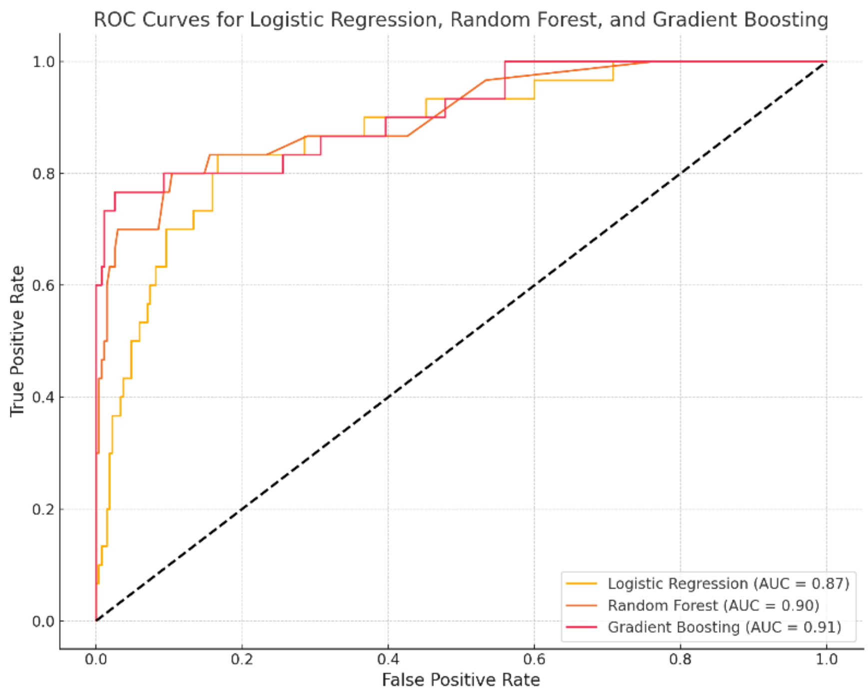

Figure 2. ROC Curve Plot

ROC curve indicates the random forecast and gradient boosting have similar overall performance, while logistics regression clearly lags. Random forest and gradient boosting has some tradeoff between level of precision and recall, but it's overall at the same level. Gradient boosting, by AUC, is the slight winner out of the two models.

Figure 3. Shap Value for Key Features

The shap value shows the most important features to determine bankruptcy. The top features include leverage, liquidity, growth, and cash on hand, which are common indicators of financial health of a company. For instance, companies are more likely to go bankrupt if it has a high level of debt and low level of cash on hand. Meanwhile, stock price variation and volatility are also key features, since the market will react if a firm shows negative signs. Since the prediction period does involve COVID-19 years, the feature does show some importance, but the impact is very minimal. Shap values from logistic regression and random forest shows most similarly results. Even though different models have different performance, they mostly use the same feature to predict bankruptcy. Figure 4 provides breakdown of false positives for the gradient boosting models.

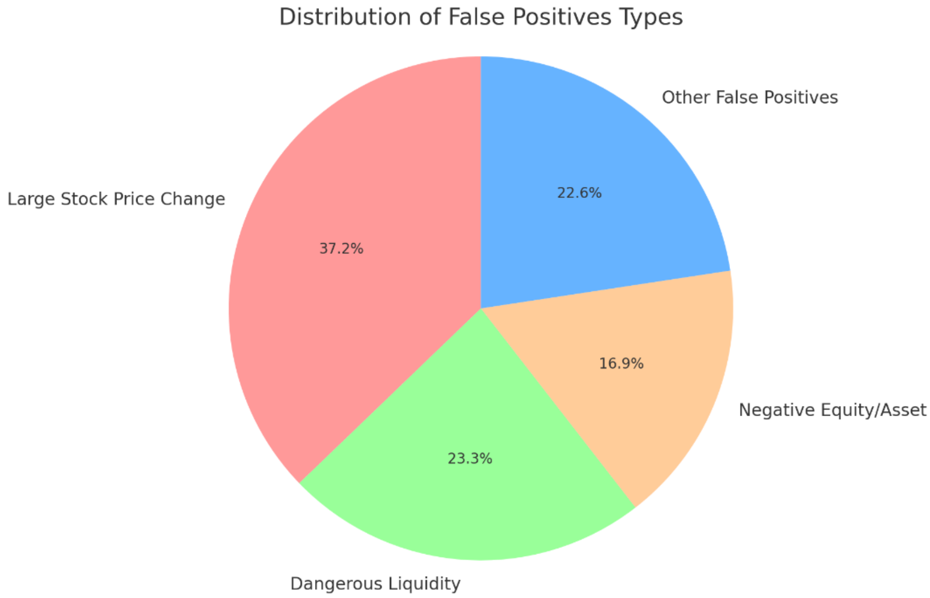

Figure 4. Analysis of False Positives

Figure 4 shows that the vast majority of false positives do correlate with actual financial risks within the firm at the time of evaluation. Among them, about 37.2% of them are flagged due to excessive high price change (drop) in the previous one-year period. Typically, a near bankruptcy firm will first crash on the stock market. Then, about 23.3% are due to low liquidity and high leverage while 16.9% of them are related to prolonged negative asset/equity values, which is very uncommon in the financial statements. Only 22.6% of the false positives cannot be explained by common signs of financial risks.

5. Conclusion

To sum up, this paper evaluates machine learning applications in the financial risk management context through fitting an imbalanced dataset with three different machine learning models. Results indicate that machine learning models can predict risks efficiently, though the performance variation is large across different models. Furthermore, the models tend to generate a large number of false positives. Even though these false positives do correlate with risks, it might still waste resources in evaluation. Regardless, careful tunning of the model is required to get the best performance and there is no standard procedure in the financial risk management sector. Companies that utilize machine learning should prepare and dedicate adequate resources for the applications.

References

[1] Guerra, P., & Castelli, M. (2021). Machine learning applied to banking supervision a literature review. [J]Risks, 9(7), 136.

[2] Leo, M., Sharma, S., & Maddulety, K. (2019). Machine learning in banking risk management:[J] A literature review. Risks, 7(1), 29.

[3] Zhang, R., Cheng, Y., Wang, L., Sang, N., & Xu, J. (2023). Efficient Bank Fraud Detection with Machine Learning.[J] Journal of Computational Methods in Engineering Applications, 1-10.

[4] Cheng, Y., Yang, Q., Wang, L., Xiang, A., & Zhang, J. (2024). Research on Credit Risk Early Warning Model of Commercial Banks Based on Neural Network Algorithm. Financial Engineering and Risk Management (2024) Vol. 7: 11-19. [OL]DOI: http://dx.doi.org/10.23977/ferm.2024.070402.

[5] Wang, L., Cheng, Y., Xiang, A., Zhang, J., & Yang, H. (2024). Application of Natural Language Processing in Fi-

nancial Risk Detection. Financial Engineering and Risk Management (2024) Vol. 7: 1-10.[OL] DOI: http://dx.doi.org/10.23977/ferm.2024.070401.
[6] Li, S., Dong, X., Ma, D., Dang, B., Zang, H., & Gong, Y. (2024). Utilizing the lightgbm algorithm for operator user credit assessment research.[J] Applied and Computational Engineering,75,36-47.
[7] Dang, B., Zhao, W., Li, Y., Ma, D., Yu, Q., & Zhu, E. Y. (2024). Real-Time pill identification for the visually impaired using deep learning.[M]arXiv preprint arXiv:2405.05983.
[8] Dang, B., Ma, D., Li, S., Qi, Z., & Zhu, E. Y. (2024) Deep Learning-Based Snore Sound Analysis for the Detection of Night-time Breathing Disorders.[J]Applied and Computational Engineering,76,109-114.