Exploration of Machine Learning Applications in Systemic Financial Risk Prediction and Management

Keyu Yao

wellington589125@gmail.com

University of Sheffield, Sheffield, United Kingdom

Abstract: In this multidisciplinary study, we explore the transformative impact of machine learning (ML) technologies in financial research. Our objective is to understand how supervised and unsupervised learning methods can be applied to tasks such as fraud detection, asset price forecasting, financial risk assessment, and the development of early warning systems for systemic financial risk. We adopt a variety of algorithms—including Backpropagation Neural Networks, Bayesian Networks, Long Short-Term Memory (LSTM) networks, Support Vector Machines (SVM), Random Forest, and XGBoost—and evaluate their ability to analyze and predict outcomes from complex financial data. Our methodology entails a comparative analysis of ML techniques against traditional statistical methods, particularly in their handling of imbalanced datasets. Our findings reveal that ML models, especially those employing ensemble and deep learning approaches, significantly enhance predictive accuracy and are superior in detecting early signs of systemic financial risks. The study demonstrates the profound advantages of ML over conventional models in constructing robust prediction models. We conclude that machine learning represents a critical advancement in the analysis of financial data, offering substantial benefits to the field of financial research and decision-making processes.

Keywords: Financial Risk Management, Machine Learning Techniques, Transparency, Systemic Risk Warning, Deep Learning

1 Introduction

Machine learning has become a transformative force in finance, streamlining complex data analysis and enhancing decision-making. This review covers how it's applied in finance, highlighting its role in fraud detection, asset forecasting, and risk assessment. Neural networks excel in detecting fraud with accuracy above 95%, while ensemble methods like Random Forest improve asset price predictions. For risk, algorithms such as XGBoost and LSTMs effectively predict credit defaults. Unsupervised learning also reveals hidden patterns in financial data, essential for risk categorization and anomaly detection. The field tackles imbalanced datasets, crucial in financial risk prediction, using combined techniques to enhance minority class predictions and prevent model bias. This synthesis points to machine learning's potential in developing early warning systems for financial risks, where models like Random Forest stand out for their efficiency with complex datasets. This literature review sets the stage for future innovation in the financial application of machine learning [1].

MSIEID 2023, December 08-10, Guangzhou, People's Republic of China

DOI 10.4108/eai.8-12-2023.2344705

2 Literature Review

Machine learning techniques, categorized into supervised and unsupervised learning, play a crucial role in financial analysis, particularly in supervised tasks such as fraud detection, asset pricing, and risk assessment. For fraud detection, Backpropagation Neural Networks and Bayesian Networks have been prominent, with the former often delivering greater speed and accuracy. Long Short-Term Memory (LSTM) models, and Support Vector Machines (SVMs) with over 95% accuracy, have also shown efficacy in this area. Regarding asset price forecasting, Random Forests standout, with ensemble methods improving their accuracy. In investment, machine learning predictions have surpassed traditional strategies. For financial risk assessment, logistic regression, decision trees, Random Forest, and XGBoost are notable, especially XGBoost and Random Forest for their accuracy in detecting risks like illegal fundraising in P2P lending. LSTMs have also been effective in predicting credit bond defaults, demonstrating high accuracy. These machine learning applications extend to systemic financial risk models, underscoring their broad utility in the finance sector [2].

3 The meaning of machine learning

Machine learning (ML) is a cornerstone of artificial intelligence, propelling the big data era forward. It differs from traditional data processing by using algorithms that allow for autonomous pattern recognition and model building without explicit programming. ML is divided into supervised learning, which uses labeled historical data to predict outcomes, and unsupervised learning, which identifies patterns in unlabeled data. Semi-supervised learning combines these approaches for partially labeled datasets, while reinforcement learning focuses on decision-making through trial and error. Advances in technology are driving ML's rapid development, with deep learning showcasing advanced capabilities in various fields, from image recognition to language processing. This growth is enhancing computers' problem-solving abilities and marking a new phase in AI's evolution. Figure 1 presents a typical framework for how machine learning operates, highlighting its key differences from traditional data processing methods [3].

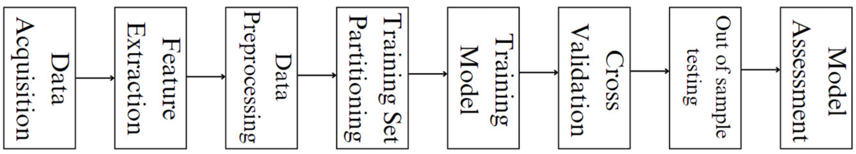

Figure 1. Framework for how machine learning operates

4 Machine learning analysis methods

Machine learning, as a core component of artificial intelligence, plays a crucial role in handling and analyzing vast amounts of data. Algorithms in this field are designed to learn and extract information from sample data, achieving predictive or classification goals through extensive iterative learning and training processes [12]. A typical machine learning workflow framework includes data collection and preprocessing, feature selection, model selection and training, and finally, testing and evaluation. Through this process, computers can learn similarly to the human brain or emulate human learning to automatically analyze historical data, discover underlying patterns, and provide insights for future decision-making and analysis. Based on different tasks and objectives, machine learning algorithms can be broadly categorized into three main types: regression algorithms, classification algorithms, and clustering algorithms [4].

4.1 Regression algorithms

Regression algorithms are used to model relationships between variables. They learn a function model based on a training dataset, allowing us to understand the relationship between independent and dependent variables and make predictions based on this relationship. For example, as shown in Figure 2, the LASSO regression model has been successfully applied to predict the impact of credit size on credit risk.

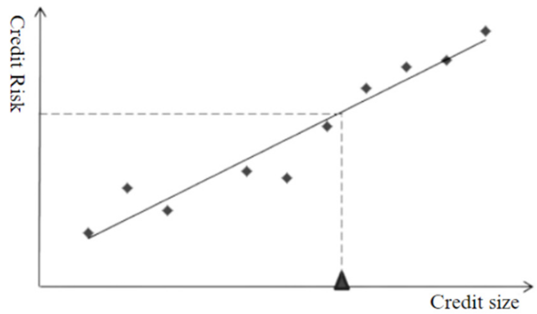

Figure 2. Classification analysis of credit risk

4.2 Classification algorithms

Classification algorithms are used to assign input data into predefined categories and are a form of supervised learning. In financial risk management, classification algorithms can be used for tasks such as financial crisis early warning and credit rating. Decision trees, logistic regression, and support vector machines are some commonly used algorithms in this domain [5]. For instance, in predicting financial distress in companies, research has shown that logistic regression models can achieve using lender age to predict credit risk, as shown in Figure 3.

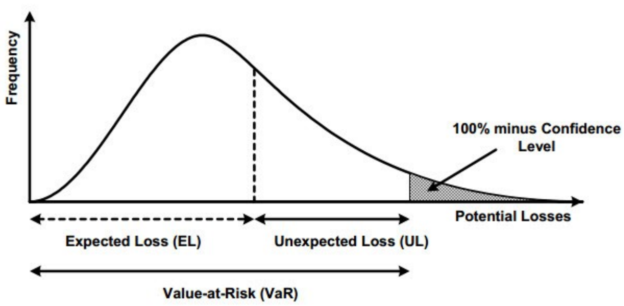

Figure 3. Classification analysis of credit risk

4.3 Clustering algorithms

Clustering algorithms are unsupervised learning methods that do not rely on predefined categories but attempt to group data points in a dataset that naturally cluster together. Data points within the same group should exhibit high similarity, while those in different groups should have significant differences. Popular clustering methods include K-means clustering, hierarchical clustering, and linear discriminant analysis. For example, as shown in Figure 4, clustering methods have been used to categorize financial risks in publicly traded companies, aiding businesses in identifying and addressing potential financial risks [6].

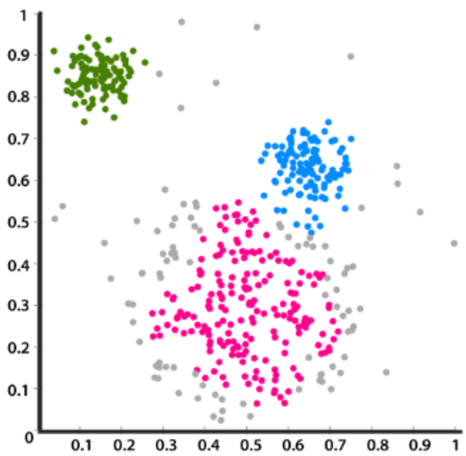

Figure 4. Clustering Analysis

5 Applications of Machine Learning in Financial Risk Prediction

5.1 Machine Learning for Imbalanced Data Set Processing

In the domain of financial risk prediction, datasets are often imbalanced, meaning the instances of financial risk are much fewer than those of normal circumstances. Training models directly on such datasets can lead to a bias towards the majority class, reducing the predictive accuracy for the minority class, which represents the occurrence of risk. To address this issue, researchers have employed methods at the data level, such as resampling techniques, as well as algorithmic-level techniques like cost-sensitive learning and ensemble learning, to balance datasets and enhance model performance. Resampling techniques include undersampling, oversampling, and hybrid sampling methods. At the algorithmic level, ensemble learning improves generalization by combining multiple models, and cost-sensitive learning improves traditional classifiers by emphasizing the cost of different types of misclassification [7].

5.2 Machine Learning in Model Construction

The application of machine learning methods in the construction of financial risk prediction models involves two crucial stages: feature selection and model establishment. Feature selection efficiently eliminates redundant features to prevent model overfitting and includes methods such as filter, wrapper, and embedded approaches [8]. The constructed models generally fall into two categories: single classifiers and ensemble classifiers, the latter of which enhances prediction performance through the combination of multiple base classifiers, such as Bagging and Boosting methods. However, with increased model complexity, interpretability can become an issue. To improve interpretability, researchers have explored intrinsic interpretation methods and model-agnostic interpretative approaches. Overall, machine learning models demonstrate superior predictive performance over traditional statistical methods and provide theoretical support for financial decision-making [9].

6 Systemic financial risk early warning model

6.1 BP neural network algorithm model

A BP (Backpropagation) neural network is a type of fully connected feedforward neural network that utilizes the backpropagation algorithm to update its parameters to minimize output error. The network consists of an input layer, one or more hidden layers, and an output layer, with each layer composed of multiple neurons. Neurons combine inputs using weights and apply a nonlinear activation function. As shown in Figure 5, the training process is supervised, where weights are repeatedly adjusted through a method like gradient descent to optimize a loss function (such as mean squared error or cross-entropy loss), until the desired outcome is achieved or a specified number of training iterations has been completed [10].

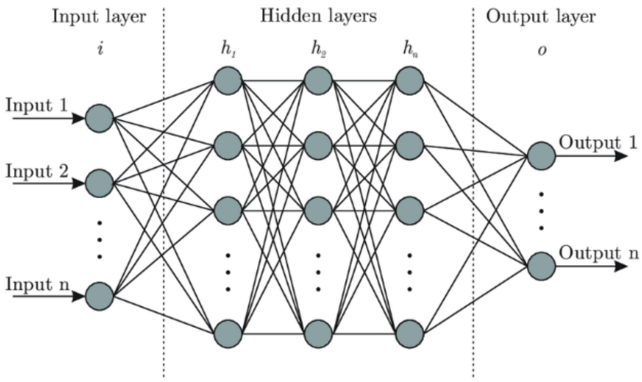

Figure 5. BP neural network model structure diagram

6.2 Analysis of Empirical Results of BP Neural Network Model

We trained and learned a BP neural network model on a dataset ranging from the first quarter of 2002 to the fourth quarter of 2020. Through this process, we obtained predicted results of systemic financial risk status from the second quarter of 2002 to the first quarter of 2021. To ensure the accuracy of the model's predictions, we further tested the model's validity before predicting the systemic financial risk status for each quarter of 2020 [11]. This validation used data from the first quarter of 2020 to the third quarter of 2021 as the validation set and employed the trained BP neural network model to test the systemic financial risk status from the first quarter of 2021 to the fourth quarter of 2021. Results are shown in table 1.

Table 1. Financial risk verification results of BPNN model (2021Q1-2021Q4)

Time Expected	Output	Expected Risk Status	Verification Output	Verification Risk Status
Q1 2021	000100	Cautious	000100	Cautious
Q2 2021	000100	Cautious	000100	Cautious
Q3 2021	000100	Cautious	000100	Cautious
Q4 2021	000100	Cautious	000100	Cautious

It can be seen that the actual output results of the model for the four quarters of the validation set sample match the expected results

7 Conclusion

In conclusion, the study vividly illustrates the profound impact of machine learning technologies on the financial sector. By leveraging both supervised and unsupervised learning strategies, machine learning algorithms such as Neural Networks, SVMs, Random Forest, and XGBoost have significantly advanced financial research capabilities. These algorithms provide sophisticated tools for fraud detection, asset price forecasting, and financial risk assessment, outperforming traditional statistical methods in precision and reliability. The exploration of machine learning in dealing with imbalanced datasets and constructing interpretable and robust prediction models has shown promising results. Additionally, the potential of these technologies in developing systemic financial risk early warning systems indicates a major shift in financial data analysis and decision-making processes. As machine learning and deep learning continue to evolve, their applications in finance are expected to become even more pervasive and integral, paving the way for more informed and accurate financial decision-making and risk management in the future.

References

[1] Henrique, Bruno Miranda, Vinicius Amorim Sobreiro, and Herbert Kimura. "Literature review: Machine learning techniques applied to financial market prediction." Expert Systems with Applications 124 (2019): 226-251.

[2] Sadgali, Imane, Nawal Sael, and Faouzia Benabbou. "Performance of machine learning techniques in the detection of financial frauds." Procedia computer science 148 (2019): 45-54.

[3] Bhatore, Siddharth, Lalit Mohan, and Y. Raghu Reddy. "Machine learning techniques for credit risk evaluation: a systematic literature review." Journal of Banking and Financial Technology 4 (2020): 111-138.

[4] Kou, Gang, et al. "Machine learning methods for systemic risk analysis in financial sectors." (2019).

[5] Rundo, Francesco, et al. "Machine learning for quantitative finance applications: A survey." Applied Sciences 9.24 (2019): 5574.

[6] Vijh, Mehar, et al. "Stock closing price prediction using machine learning techniques." Procedia computer science 167 (2020): 599-606.

[7] Vachhani, Hrishikesh, et al. "Machine learning based stock market analysis: A short survey." Innovative Data Communication Technologies and Application: ICIDCA 2019. Springer International Publishing, 2020.

[8] Rasekhschaffe, Keywan Christian, and Robert C. Jones. "Machine learning for stock selection." Financial Analysts Journal 75.3 (2019): 70-88.
[9] Buduma, Nithin, Nikhil Buduma, and Joe Papa. Fundamentals of deep learning. " O'Reilly Media, Inc.", 2022.
[10] Thompson, Neil C., et al. "The computational limits of deep learning." arXiv preprint arXiv:2007.05558 (2020).
[11] Van der Laak, Jeroen, Geert Litjens, and Francesco Ciompi. "Deep learning in histopathology: the path to the clinic." Nature medicine 27.5 (2021): 775-784.