Application Analysis of Machine Learning Models in Credit Risk

Shiyi Liu*

University of Business School, University of Nottingham, Ningbo, China

* Corresponding Author Email: biysl45@nottingham.edu.com

Abstract. Traditional credit risk assessment models frequently depend on historical data and predefined rules, a limitation that may fail to fully encompass the intricate and dynamic characteristics of credit risk. With the rise of machine learning (ML), credit risk can be assessed more easily according to historical data, and it offers a more dynamic and data-driven approach to credit risk assessment. This essay will provide an overview of significant research advancements made in the last five years. It demonstrates consumers and financial institutions have different preferred ML models. Moreover, using different criteria such as AUC and Brier to evaluate accuracy yields different results. And there is a greater prospect for deep learning. Overall, the advent of machine learning has revolutionised the field of credit risk modelling, offering new insights and methodologies for predictive analytics. These advancements utilise statistical, machine learning, and deep learning methods to tackle credit risk issues. In addition, this essay will explore several obstacles and forecast future trends.

Keywords: Credit risk, Machine learning, Credit scoring, Deep learning.

1. Introduction

Credit risk poses a persistent challenge in the financial industry, characterized by the potential for a counterparty to default on its obligations as per the agreed terms, leading to financial losses for the lender. Minimizing these obligations is key to managing risk for financial organizations, as the Basel II New Capital Accord sets new standards for bank capital adequacy, and optimizing capital allocation is of great significance. Therefore, the accurate and effective assessment of credit risk models is necessary for every financial institution to mitigate potential losses.

Credit scoring models have undergone a lot of evolution and improvement. The traditional qualitative analysis 5C model has confirmed the disadvantage of handling mass data and making subjective judgments, but this resulted in treating borrowers unfairly with several subjective factors [1]. In contrast, quantitative analysis seems to be more objective. Effective credit risk models draw upon huge amounts of data, typically incorporating variables such as credit scores, debt-to-income ratios, and even macroeconomic indicators. These variables are then processed through statistical techniques or machine learning algorithms to generate predictions about a borrower's financial behaviour. Thus, to better evaluate the effectiveness of models, credit scoring systems have experienced lots of optimization. Initially, analysts use basic linear regression to calculate credit scores, a method involving statistical techniques that identify the relationship between borrowers' characteristics and their likelihood of default. Over time, as computational power increased, more intricate models, such as decision trees, were utilized, offering increased precision by partitioning the data into subsets based on feature-value conditions.

The evolution didn't stop there. The introduction of ensemble techniques like random forests and gradient boosting further refined credit assessments. These methods combine predictions from multiple trees to improve accuracy and robustness and avoid overfitting. Recently, neural networks and deep learning algorithms have been used for credit scoring. Their ability to capture nonlinear patterns and interactions between variables is unrivalled but at the cost of requiring large amounts of data and computational resources.

Each step of this evolution necessitated the development of better data processing and model evaluation metrics. From confusion matrices to ROC curves and AUC metrics, analysts have a suite of tools to evaluate the performance of credit scoring models. In addition, as machine learning models become more complicated, the need for interpretability and fairness has led to the creation of techniques to interpret model predictions and ensure that algorithms do not propagate bias in the credit decision-making process.

By doing so, lenders are equipped with a quantitative basis upon which they can devise their lending strategies, ensuring that loans are issued in a manner that aligns with the institution's risk tolerance and regulatory requirements. Moreover, the utility of these models extends beyond new credit issuance. They are deeply integrated into the risk management frameworks of banks and other lending institutions where they support ongoing monitoring of credit portfolios. This continuous oversight enables the early identification of deteriorating creditworthiness among existing borrowers, allowing for timely interventions such as loan restructuring or the setting aside of additional capital reserves to buffer potential losses.

Credit risk models do not only protect lenders but, by promoting responsible lending practices, they also shield the broader economy from the ripple effects of credit defaults. Therefore, the development and refinement of these models are of paramount importance, requiring rigorous testing and validation to ensure their accuracy and reliability over time.

2. Literature Review

Khandani et al. applied machine learning techniques to create a nonlinear and nonparametric predictive model for consumer credit risk. Their analysis of the cost-benefit implications of using machine learning predictions indicated potential cost savings ranging from 6% to 25% of total losses [2]. Additionally, Kruppa et al. established a comprehensive framework in 2013 for evaluating the credit risk of individual consumers through the utilization of machine learning methods [3].

Butaru et al. extended the application of machine learning beyond consumer credit risk prediction to the credit card industry. By analyzing account-level credit card data from six major commercial banks, they leveraged machine-learning techniques to forecast delinquencies. Their approach involved integrating various data sources such as consumer transaction lines, credit bureau information, and macroeconomic variables. This approach not only enhances the prediction of delinquencies but also allows for the examination and comparison of risk management practices and delinquency drivers across different banks. As a result, random forests effectively measure loss probability and credit risk [4]. Particularly post the 2007-2008 financial crisis, there has been heightened awareness among firms regarding the pivotal role of credit scoring systems in credit risk management. In parallel, Luo et al. examined the efficacy of credit-scoring models when applied to Credit Default Swap (CDS) datasets, suggesting ongoing research into refining credit risk management practices in response to market dynamics and financial crises [5].

In 2019, Bao et al. introduced an innovative approach to credit risk management by combining unsupervised and supervised learning techniques [6]. Unlike prior studies, they applied unsupervised learning at two stages: the consensus phase and the data aggregation phase. Their use of the self-organizing mapping (SOM) algorithm helped mitigate the risk of overfitting in constructing the consensus model. Simultaneously, SME credit risk prediction in Supply Chain Finance (SCF) emerged as a significant concern in financing decisions. Zhu et al. addressed this by combining two classical ensemble machine learning methods, Random Subspace (RS) and Multiple Boosting (Multibooting), to create an enhanced hybrid ensemble ML approach named RS-Multibooting. They aimed to improve the accuracy of credit risk prediction in this domain [7].

Overall, credit scoring (CS) stands as a pivotal tool for risk management in banking and financial institutions. Plawiak et al. introduced a novel method named Deep Genetic Hierarchical Learning Network (DGHNL) [8]. Meanwhile, Moscatelli et al. assessed various machine learning models against standard statistical methods like logistic regression to predict default risk [9]. Their findings suggested that machine learning models prioritize credit allocation to safer, larger borrowers, potentially reducing lenders' credit losses. Additionally, Bhatore et al. conducted a comprehensive literature review covering 136 papers on credit risk assessment methods and machine learning techniques published between 1993 and March 2019 [10].

3. Methodology

As shown in Fig. 1, machine learning (ML) models consist of three primary components: supervised learning, unsupervised learning, and reinforcement learning. A series of recent studies have shown good results with random forest and decision tree models, deep learning, and several ensemble methods, as described in more detail below.

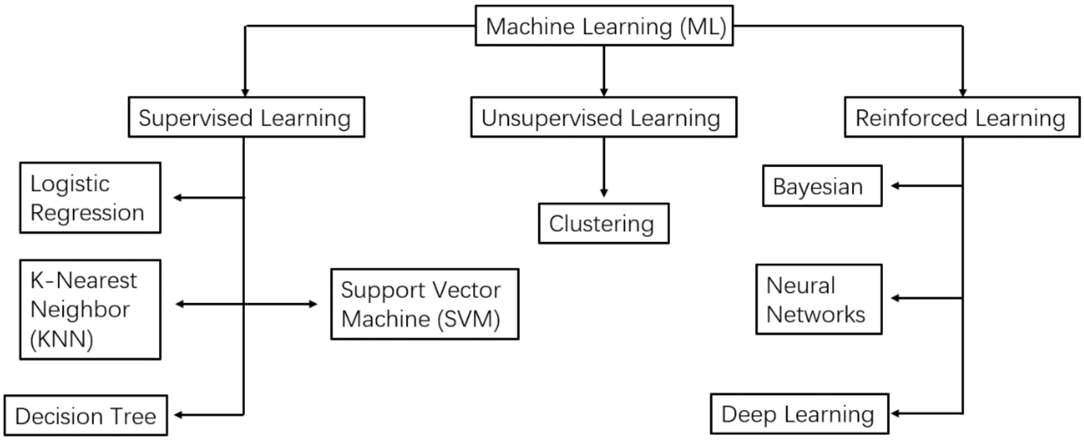

Fig. 1 Machine learning landscape (Photo credit: Original)

3.1. Random Forests and Decision Tree Models

In a study by Kruppa et al., it was shown that random forests exhibited superior performance compared to adapted logistic regression when analyzing a substantial credit-scoring dataset [3]. Similarly, Butaru et al. conducted research indicating that decision trees and random forests surpassed logistic regression in predicting credit card delinquency for out-of-sample and out-of-time scenarios. Remarkably, this advantage was notably significant, especially when considering a short time horizon [4]. Thus, random forests are an ensemble learning method, comprising multiple decision trees, which produce a collective output that typically has better predictive accuracy and robustness than individual trees. Decision trees, the building blocks of random forests, are simple to understand and interpret. They split the data into branches to arrive at a decision or prediction, which makes them particularly useful for classification tasks like credit risk evaluation.

In credit risk assessment, decision trees can handle both categorical variables (e.g., marital status, employment type) and continuous variables (e.g., income, loan amount). When building a decision tree for credit risk, the algorithm selects the variable and the split-point that result in the most significant differentiation between those who default and those who do not, based on measures like Gini impurity or information gain.

Random forests take this a step further by generating numerous decision trees, each trained on a random subset of the dataset (with replacement, called bootstrap aggregation or bagging) and using a random subset of features for each segmentation. This randomness helps to avoid overfitting, which is a common problem with individual decision trees, as overfitting causes the model to fit too closely to the training data, thus capturing noise along with the signal. The random forest combines the predictions from all the individual trees, usually by voting for classification tasks, which smooths out the predictions and gives a more accurate assessment of credit risk.

Through these techniques, random forests and decision trees provide valuable tools for financial institutions to determine the likelihood of a borrower defaulting on a loan. They are capable of accommodating various types of data, and their inherent interpretability makes them particularly attractive for making and explaining credit decisions.

3.2. Deep Learning Approaches

Deep learning models, such as the Deep Belief Network (DBN) shown in Fig. 2, utilise multi-layer non-linear information processing. For example, DBNs consist of multiple layers of random latent variables. They consist of simple, unsupervised networks (e.g., restricted Boltzmann machines) that are "stacked" on top of each other. Each layer communicates with both the previous and subsequent layers, forming a deep architecture that can learn to represent data hierarchically. Luo et, al. have proved that DBN is superior to other algorithms in several variables, like accuracy, AUC, and misclassification [5].

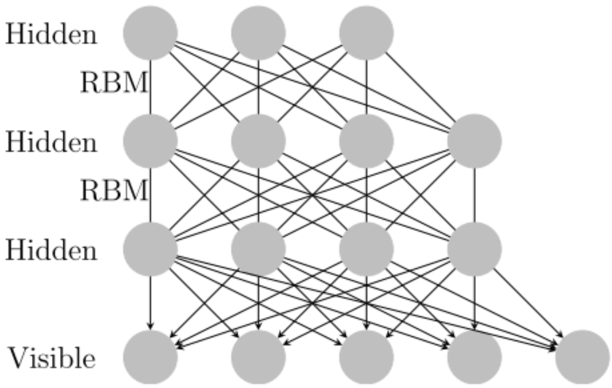

Fig. 2 DBN [5]

Through the layers of neurons, deep learning models can learn to discern the non-linear relationships and hidden structures within these features. An adept deep learning system may reveal that a combination of a high credit utilization ratio and a series of recent credit inquiries correlates strongly with a higher risk of default. This ability to forecast credit risk with high accuracy is invaluable to banks and financial institutions that seek to minimize the risks associated with lending while also providing fair and accessible credit options to consumers.

3.3. Ensemble Methods and Hybrid Approaches

Ensemble and hybrid approaches in machine learning bring together a diverse array of algorithms to tackle the complexity of credit risk assessment. The rationale behind these methods is that by harnessing the strengths of various algorithms and mitigating their weaknesses, one can create a more robust and accurate predictive model.

For hybrid models, which are a specialization within ensemble techniques, the process often involves a blending of different machine learning paradigms such as supervised and unsupervised learning. In terms of credit risk, this could mean using unsupervised methods to segment consumers into various risk categories and then applying different supervised learning algorithms to predict the likelihood of default within each segment. For example, Plawiak et al. introduced a new Deep Genetic Hierarchical Network of Learners (DGHNL), which consists of different types of learners, and achieved an accuracy of 94.60% when used for credit score prediction [8].

4. Challenges and Future Directions

4.1. Overcoming Limitations of Current Models

Despite their advantages, ML (Machine Learning) models encounter various limitations that can hinder their effectiveness. The major defect in their experiments is that they lack vast sample datasets, and the impact of macroeconomic and other factors is not considered. Another limiting factor is overfitting, arising when a model captures excessive detail and noise from the training data, consequently impairing its performance on unseen data. This is often a result of a model being too complex relative to the amount of training data or a lack of regularization. Regularization methods such as L1 and L2 regularization, pruning, or cross-validation can be used to prevent overfitting by making the model simpler or penalizing complex models.

Lastly, interpretability challenges arise because some ML models, particularly complex ones like deep neural networks or ensemble methods, act as 'black boxes' with their decision-making process being difficult to understand and explain. This is problematic in domains where interpretability is crucial, such as in healthcare or finance. To address this, techniques like feature importance scores, model-agnostic methods, or simpler models that are inherently more interpretable are leveraged. Overall, while ML models have transformative potential, addressing these limitations is crucial for creating reliable and trustworthy systems.

4.2. Trends

The potential of Machine Learning (ML) is vast, particularly when it comes to the realm of finance. One crucial application is in the augmentation of individual credit assessments.ML algorithms analyze a variety of data points that traditional assessments might overlook, such as shopping patterns, social media activity, and even mobile phone usage to gauge a person's creditworthiness with greater precision.

Therefore, the value of ML in the financial sector is multifaceted which improves individual credit assessments and expands the capabilities to predict and navigate through systemic risks. Such advanced analytics are becoming an integral part of maintaining financial stability in an increasingly complex and interconnected world. It is not just an upgrade for financial processes.

5. Conclusion

Machine learning's integration into credit risk modelling has revolutionized the way financial institutions assess and manage risks associated with lending. Its ability to churn through vast datasets, identify complex patterns, and predict outcomes with a high degree of accuracy allows for a more nuanced understanding of creditworthiness. The predictive analytics derived from machine learning algorithms can facilitate decision-making processes, improve the speed and efficiency of credit assessments, and potentially lower the cost of credit for consumers.

Through plenty of literature, many different machine learning algorithms have demonstrated the capability to efficiently handle credit risk. But there is still much to be improved. Deep learning is expected to continue to make deep inroads as it is proven to work better than other tools for machine learning. In addition, accurate datasets are a big aspect to consider. The adage "garbage in, garbage out" is particularly relevant in terms of machine learning. The effectiveness of credit risk models hinges entirely on the quality of the data used for their training. Inaccurate, incomplete, or outdated data can skew predictions and lead to incorrect credit risk assessments. Financial institutions need to institute rigorous data management protocols to uphold data accuracy, comprehensiveness, and currency.

References

[1] Dastile X, Celik T, Potsane M. Statistical and machine learning models in credit scoring: A systematic literature survey. Applied soft computing, 2020, 91:106263.

[2] Khandani A E, Kim A J, Lo A W. Consumer credit-risk models via machine-learning algorithms. Journal of Banking & Finance, 2010, 34(11): 2767-2787.

[3] Kruppa J, Schwarz A, Arminger G, Ziegler A. Consumer credit risk: Individual probability estimates using machine learning. Expert Systems with Applications,2013, 40(13): 5125-5131.

[4] Butaru F, Chen Q Q, Clark B, Das S, Lo A W, Siddique A. Risk and risk management in the credit card industry. Journal of Banking and Finance, 2016, 72: 218-239.

[5] Luo C C, Wu D S, Wu D X. A deep learning approach for credit scoring using credit default swaps. Engineering Applications of Artificial Intelligence, 2017, 65: 465-470.

[6] Wang B, Ning L J, Kong Y. Integration of unsupervised and supervised machine learning algorithms for credit risk assessment. Expert Systems With Applications, 2019, 128: 301-315.

[7] Zhu Y , Zhou L, Xie C, Wang G J, Nguyen T V. Forecasting SMEs' credit risk in supply chain finance with an enhanced hybrid ensemble machine learning approach. International Journal of Production Economics, 2019, 211: 22-33.

[8] Plawiak P, Abdar M, Plawiak J, Makarenkov V, Acharya U R. DGHNL: A new deep genetic hierarchical network of learners for prediction of credit scoring. Information Sciences, 2020, 516: 401-418.

[9] Moscatelli M, Parlapiano F, Narizzano S, Viggiano G. Corporate default forecasting with machine learning. Expert Systems with Applications, 2020, 161: 113567.

[10] Bhatore S, Mohan L, Reddy Y R. Machine learning techniques for credit risk evaluation: a systematic literature review. Journal of Banking and Financial Technology, 2020, 4: 111-138.