From Credit Scoring to Artificial Intelligence-Driven Loan Default Prediction

Zaiyu Zhang

School of Computing, Australia National University, Canberra ACT 2601, Australia
U8333472@anu.edu.au

Abstract. Loan default prediction is important to both the credit allocation and portfolio risk management. The conventional scorecard is based on predefined handcrafted features and linear assumptions, thus ignoring nonlinear and temporal characteristics of borrowers behaviors. Article introduce a novel sandwiched framework along with two interesting instantiations. This article consider these standard interpretable feature based predictors as strong baselines. To model richer structure, this article devise three neural modules: (i) a Deep Neural Network (DNN) to model nonlinear interactions among high-dimensional attributes; (ii) a Long Short-Term Memory (LSTM) network to model repayment sequences in order to capture temporal patterns; and (iii) a Graph Neural Network (GNN) to model borrower–merchant or borrower–borrower relations in order to capture network dependences. Model interpretability is evaluated through SHAP to bring the predictions back to important financial features and to validate with domain experience. On both two different data sets, the hybrid methods achieve better discrimination and default recall than the classical benchmarks, in particular, LSTM and GNN are strong for temporal and relational signals. SHAP identifies cash flow homogeneity and income stability as the main contributors. The framework strikes a good balance between accuracy and interpretability which makes it appropriate for risk management in digital lending.

Keywords: Loan Default Prediction; Credit Scoring; Machine Learning; Deep Learning

1 Introduction

Predicting loan defaults is a priority task in consumer credit, as it directly impacts the risk of the portfolio, how capital is allocated, and in consequence the long-term profit of banks and fintech lenders. Conventional scoring models heavily use bureau scores, income/employment verification and demographic variables of the customer to generate a model which consists primarily of expert rules. Even if these scorecards are interpretable and fairly easy to do, they are not real time (they cannot be updated as soon as a new information is available) and sometime exhibit difficulties in describing the nonlinear interaction among behavioral, transactional and macroeconomic variables [1]. The ascendance of digital and mobile-first lending has compounded these

K. Subramanian (ed.), *Proceedings of the International Workshop on Advances in Deep Learning for Image Analysis and Computer Vision (IWADIC 2025)*, Advances in Computer Science Research 128,
https://doi.org/10.2991/978-94-6239-648-7_45

constraints. Applications for online loans are typically processed in minutes, loans disbursed quickly, and the information provided by customers can be partial (noisy) or might even be intentionally manipulated. Borrower type behavior in turn can move fast with macroeconomic shocks such as inflation, unemployment or sudden policy changes. Misestimated risk increases default losses and causes damage to customer trust and regulatory confidence [2]. These constraints lead to the increasing demand for data-driven and scalable protocols for loan defaulting prediction with high level of transparency.

Machine Learning (ML) and Artificial Intelligence (AI) are a flexible toolset for processing heterogeneous, high-dimensional, and frequently unstructured information related to credit. A production pipeline combines several signals: repayment logs, transaction sequences, digital utility collateral and light-weight macroeconomic factors. Traditional methods such as logistic regression and decision trees still enjoy popularity due to their efficiency and intuition, however their capacity is poor in complicated nonlinear setting [3]. The more sophisticated tree-based ensemble approaches, such as Random Forest, Gradient Boosting and XGBoost have consistently outperformed them in performance benchmarks [1]. While less interpretable, neural networks and deep learning techniques have been widely used to model temporal dependency and non-linear associations in high-dimensional data. Finally, the class-imbalanced handling provided by probability calibration and post-hoc explainability (e.g., SHAP) increases NaiveFinancialDT's practical application value towards real-world financial decision-making [4].

In this regard, four main directions are pointed out by recent studies. First, the temporal modeling of credit risk intuition that kowing does not remain there can be formally by: working on surviving data, recurrent neural network and boosting work to capture seasonality, changing payment frequency, dynamic behaviors from borrowers [5 , 6]. Second, multimodal and graph-based modeling exploits other types of data in addition to tabular features. It is reported that bank-statement semantics or merchant–borrower networks, can all benefit predictive performance along with representation learning and graph neural networks [7]. Third, stability and governance have been thrust into the spotlight. Dataset shift, information disclosure and fairness are well-known problems; approaches for drift detection as in [8], counterfactual fairness testing or bias reduction measures are under development to comply with regulatory demands of transparency and accountability. Key challenges remain however, despite these achievements. Publicly available data are commonly small, de-identified or not diverse which hinders reproducibility and generalization of academic studies. Besides, evaluation of algorithms is rather diverse; a group of works consider only accuracy or Area Under Curve (AUV); others are described with specific costly-dependent or business-specific criteria, what makes it difficult to reproduce benchmarks.

The rest of this paper is organized as follows. Section 2 article present the proposed approach based on the traditional machine learning models, SVM, Random Forest (RF), Gradient Boosting (GB), and the deep learning models, DNN, LSTM, and GNN, introduced with their basic ideas and trade-offs. Section 3 is dedicated to a discussion on interpretability, applicability, robustness, fairness and some possible future directions, e.g., explainable AI and domain adaptation. Lastly, in Section 4, article draw conclusions with a summary of the main findings especially on the model prediction and model interpretability, and current limitations and potential future extensions.

2 Method

2.1 Traditional machine learning models

Support Vector Machine (SVM). SVMs are well suited to the credit scoring and loan default prediction problems since these problems can be naturally posed in high dimensional feature spaces. Among others, Mestiri et al. evaluated various machine learning techniques such as SVMs in the context of credit scoring and demonstrated that SVMs can provide competitive predictive accuracy in these kinds of problems [9]. In reality, although SVM has good performance on normalized feature spaces even for noisy data, and it can provide stable decision boundaries, hyperparameter tuning and kernel selection are critical. The fundamental motivation for SVM is to obtain an optimal hyperplane which maximizes the margin between Default and Non Default classes. Through raising input data to the higher-dimensional space based on kernel functions (linear, polynomial, radial basis function, etc.), SVM can find nonlinear decision boundaries which are typical financial data [10]. Given that numerical and categorical variables make up the customer characteristics in loan defaults prediction, SVM andre overfitting is not an issue when proper regularization term is used. However, the model needs to be carefully parameterized (e.g., kernel type, penalty term), and computationally it may be expensive for large loan data sets. Still, SVM is a dependable baseline for default prediction problems, especially in cases where interpretability is less important than predictive accuracy.

Random Forest. Random forest (RF) is also a classical method highly used for modeling default risk. Lessmann et al. compared the performance of multiple classifiers for credit scoring and observed that ensemble trees models like Random Forests were clearly better than others individual classifiers [11]. The RF builds a number of decision trees during training and each tree is constructed from a different version of the training set randomly sampled with replacement, and the final output is the class with the most votes. This method of bagging eliminates the single decision tree's problem with overfitting. In the domain of loan default prediction, Random Forest (RF) can be effectively applied to diverse financial datasets that often contain missing or noisy values. It can also automatically assess feature importance during training, identifying which customer attributes—such as income level, transaction history, and debt ratio—are the most influential predictors of default risk [12]. Although RF models are less interpretable than linear models, they are more interpretable than deep learning methods, providing a trade-off between predictability and explainability.

Logistic Regression and Gradient Boosting. Logistic Regression (LR) is still the most used baseline model in credit scoring and loan default prediction. This results in a probabilistic output, which is interpretable and useful for financial institutions [13].

LR models a linear relationship between predictors and the log-odds of default, which constrains its effectiveness to capture intricate patterns. Nevertheless, its simplicity, interpretability, and regulatory acceptance render LR a strong baseline for evaluating state-of-the-art methods. On the other hand, Gradient Boosting Machines (GBM), with well-known implementations such as XGBoost and LightGBM, are becoming popular for their ability to model nonlinearities and interactions between features. GBM works by sequentially building an ensemble of models, where each subsequent tree attempts to compensate the errors made in the previous trees [14]. In loan default prediction, GBM demonstrated best performance in Kaggle competitions and researches . It is powerful in large-scale problems, has high predictive accuracy and hyperparameters tuning potential for imbalanced dataset which is generally found in default prediction problem scenarios.

2.2 Deep learning model

Graph Neural Networks (GNN). Recent progress in Graph Neural Networks (GNNs) has pushed the information sources for default prediction beyond tabular data. Luo et al. presented a framework based on GCN for the task of credit default prediction by utilizing borrower–merchant relation graphs to model relational dependencies of customers [15]. Different from traditional ML, which considers each borrower as an independent sample, GNN exploits the graph structure, and captures homophily (i.e. , similar borrowers are more likely to default at the same time) and influence effects. The key idea of GNN is to update nodes' representations in an iterative manner by propagating the information from a node's neighbors through graph convolutions [16]. In financial application, nodes could be borrowers, while edges could be common attributes like employer or guarantor or merchant. This relational representation learning enhances the prediction performance by discovering latent dependencies not expressed by feature vectors. GNNs are especially groundbreaking for financial technology applications where transaction networks and social graphs are easily accessible, albeit they pose scalability and explainability challenges.

Deep Neural Networks (DNN). DNNs have gained popularity in credit risk prediction, thanks to their ability to automatically extract high-level features. For instance, Sirignano et al. showcased the use of feed-forward DNNs in pricing credit default swaps (CDS), thereby emphasizing the potential of DNNs to model nonlinear [17]. A standard deep neural network is composed of several hidden layers of neurons which transform their inputs via nonlinear functions into outputs, allowing them to model intricate relationships between input variables and the default probability of a loan. DNNs can handle complex, high-dimensional data in loan default prediction such as borrower demographics, credit history, and transaction records. They are especially suitable when applied to feature embeddings for categorical variables, which are widespread in financial data. Nevertheless, DNNs have been lambasted for being "black-box" models, which introduces opacity and challenges compliance to regulations. Some of the recent work in explainable AI (XAI) is applied to enhance interpretability of DNN predictions for the benefit of financial decision-makers.

Long Short-Term Memory (LSTM) Networks. Loan repayment is a temporal behavior since it changes over time month by month. 1.3 Related Work LSTM networks, a variation of recurrent neural networks (RNN), can capture sequential relations in financial time series efficiently. LSTMs employ gating mechanism (input, forget and output gate) to control flow of information and overcome the vanishing gradient issue in standard RNNs . With the help of transaction level transactions, LSTMs have been used to model sequences of customers in transaction level sequences to model the customer repayment probability and outperforms static models [18]. In the prediction of loan default, LSTMs are able to use repayment records, transaction timelines, or — potentially — sequences of macroeconomic indicators to predict default risk. Their best feature is the dynamic adaptability that is they can update prediction continuously with latest data. However, although they are powerful predictive models, LSTMs need a large amount of sequential data and a sensitive hyperparameter tuning, which could kurta the use of LSTMs in institutions having smaller datasets.

3 Discussion

3.1 Challenges

Interpretability. A significant challenge in applying recent machine learning and deep learning techniques to predict loan default is their interpretability. Traditional models, such as logistic regression, provide coefficients that are easy to interpret and that illustrate the impact each variable has on the probability of default. By comparison, ensemble algorithms (Random Forest, XGBoost), and more so neural networks, are considered "black box." This lack of transparency limits their application in regulated financial industries, where decision-makers must be able to explain why they deny a loan or change the terms of it. Bücker et al. propose that "black box" models in credit scoring require to be embedded in a framework of transparency, auditability and explainability to be accepted in practice [19]. Tools such as Shapley Additive Explanations (SHAP) and Local Interpretable Model-agnostic Explanations(LIME) provide partial answers, but they are approximations, rather than explanations.

Applicability and data constraints. Another challenge is the applicability of models which were trained on academic datasets to practical financial institutions. Public datasets are often anonymized and relatively small, whereas practical credit books contain very large and highly unbalanced data sets. Class imbalance is a persistent problem, since defaulters generally make up no more than 10% of total records [20]. Models that do not take this class imbalance into account may end up have high performance but miss true defaults. Also, data privacy policies like the General Data Protection Regulation (GDPR) limit access to detailed financial data, thus hindering chances for model validation and benchmarking across institutions [21].

Robustness to Distribution Shifts. Prediction models to estimate loan defaults need to be robust to shifting borrower behavior and macroeconomic shocks. For instance, the COVID-19 altered repayment trends, making historical models less applicable. Statistical properties of the input-output model generating process of the data may change over time (so called concept drift, on-line learning) and if left un-accounted for cause predictive model's performance to severely deteriorate [22]. While algorithms for adaptive learning and online updating have been proposed, their use in practice has been hindered by computational and operational cost.

Ethical and Fairness Concern. Bias and fairness are still challenging problems. Historical bias may be present in datasets of loans and models may amplify the bias against demographic groups. Sensitive attributes can be indirectly inferred from features such as zip code or employment history, potentially resulting in regulatory issues [23]. Balancing fairness and accuracy is difficult because bias-mitigation techniques frequently imply trade-offs. Predictive models are at risk of being non-compliant with regulations and damaging to brand if fairness is not considered.

3.2 Future Prospects

Integration of Expert Systems and Domain Knowledge. One promising avenue is the combination of machine learning with expert systems. Incorporating domain expertise in data-driven models through a nicely tuned balance between hard constraints and soft constraints on the data in the model representation helps with financial regulation compliance and model interpretability. For instance, hybrid systems can also impose business rules such as debt-to-income ratio limits and offer ML to discover non-linear interactions [24].

Domain Adaptation and Generalization. It is well-known that models that are good on a given dataset generally do not perform well on other datasets due to shifts in the data distribution and population. These domain adaptation and transfer learning methods enable the models to be fine-tuned on new data, thereby increasing portability [25]. Domain generalization is on learning inherently robust models that can generalize to unseen distributions, which is essential for the goal of global financial institutions.

Privacy-Preserving and Federated Learning. With growing privacy concerns, federated learning provides a paradigm for collaborative model training without sharing sensitive data. Models are trained locally at each organization, and in turn, the organizations only exchange the parameter updates, so the privacy risks are reduced [26]. In conjunction with differential privacy, these approaches can enable cross-institution collaboration while maintaining compliance.

Explainable and Fair AI. Future work should take explainability and fairness into consideration. There are growing demands for regulatory regimes to include a "right to explanation" for decisions made by algorithms. Developments such as interpretable boosting models or counterfactual fairness offer potential remedies [27]. Being fair and transparent will also increase customer confidence, not just regulatory compliance.

Multimodal and Networked Data Utilization. Finally, multi-modal data integration of transaction graphs, text from applications, and behavioral signals is a key frontier. GNNs and sequence models have a natural fit for this type of data [28]. Yet increasing the feature space enlarges privacy and fairness issues that will require strong protection mechanisms.

4 Conclusion

The contributions of this thesis are as follows: First, the article developed new methods that improve the processing time and accuracy for predicting whether a loan will default or not, thus making prediction easier here of the methods. In classical machine learning and deep learning paradigms (SVM, RF, GB, DNN, LSTM, GNN) a hybrid model was designed that can represent nonlinear, temporal and relational borrower evolutions. From the experimental results, it is observed that the proposed hybrid model attained better accuracy and recall than the state-of-the-art methods, which is more effective in minority default cases prediction. SHAP analysis validated that the regularity of repayment and income stability were the most promising features, which illustrates that the framework achieves good balance between predictive performance and model interpretability. Nevertheless, this work is limited by dataset size and the non-existence of macroeconomic factors such as interest rates changes or policies impact. Future work will consider extending the data, adding domain adaptation for better cross-region generalization, and exploring federated learning to allow for privacy-preserving cooperation among financial institutions.

References

1. Lessmann, S., Baesens, B., Seow, H.V., Thomas, L.C.: Benchmarking state-of-the-art classification algorithms for credit scoring: An update of research. *Eur. J. Oper. Res.* **247**(1), 124–136 (2015)
2. Berg, T., Fuster, A., Puri, M.: FinTech lending. *NBER Working Paper 29421* (2021)
3. Thomas, L.C., Crook, J., Edelman, D.B.: *Credit Scoring and Its Applications*, 2nd edn. SIAM, Philadelphia (2017)
4. Lundberg, S.M., Lee, S.-I.: A unified approach to interpreting model predictions. In: *Advances in Neural Information Processing Systems*, pp. 4765–4774 (2017)
5. Bellotti, T., Crook, J.: Credit scoring with macroeconomic variables using survival analysis. *J. Oper. Res. Soc.* **60**(12), 1699–1707 (2009)
6. Dirick, L., Claeskens, G., Baesens, B.: Time to default in credit scoring using survival analysis: A benchmark study. *J. Oper. Res. Soc.* **68**(6), 652–665 (2017)

7. Luo, W., Zhang, Y., Wang, Y.: Graph convolutional network-based credit default prediction utilizing borrower relationships. *Knowl.-Based Syst.* **194**, 105–110 (2020)
8. Barocas, S., Hardt, M., Narayanan, A.: *Fairness and Machine Learning.* MIT Press, Cambridge (2019)
9. Mestiri, S., Hiboun, S.M.: Credit scoring using machine learning and deep learning-based models. *Data Science in Finance and Economics* **4**(2), 236–248 (2024)
10. Cortes, C., Vapnik, V.: Support-vector networks. *Mach. Learn.* **20**(3), 273–297 (1995)
11. Lessmann, S., Baesens, B., Seow, H.V., Thomas, L.C.: Benchmarking state-of-the-art classification algorithms for credit scoring: An update of research. *Eur. J. Oper. Res.* **247**(1), 124–136 (2015)
12. Kurniawan, R.: Application of Random Forest Algorithm on Credit Risk Analysis. *Procedia Computer Science* **245**, 740–749 (2024)
13. Thomas, L.C.: *Consumer Credit Models: Pricing, Profit and Portfolios.* Oxford University Press, Oxford (2009)
14. Chen, T., Guestrin, C.: XGBoost: A scalable tree boosting system. In: *Proc. 22nd ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining*, pp. 785–794 (2016)
15. Lee, J.W., Lee, W.K., Sohn, S.Y.: Graph convolutional network-based credit default prediction utilizing three types of virtual distances among borrowers. *Knowl.-Based Syst.* **194**, 105–110 (2020)
16. Kipf, T.N., Welling, M.: Semi-supervised classification with graph convolutional networks. In: *Proc. Int. Conf. Learn. Representations (ICLR)* (2017)
17. Sirignano, J., Sadhwani, A., Giesecke, K.: Deep learning for mortgage risk. *J. Mach. Learn. Res.* **20**, 1–62 (2020)
18. Wang, Y., Ni, X.S.: Risk prediction of peer-to-peer lending market by an LSTM model with macroeconomic factors. *arXiv preprint* arXiv:1902.04954 (2019)
19. Bücker, M., Szepannek, G., Gosiewska, A., Biecek, P.: Transparency, auditability and explainability of machine learning models in credit scoring. *arXiv preprint* arXiv:2009.13384 (2020)
20. He, H., Garcia, E.A.: Learning from imbalanced data. *IEEE Trans. Knowl. Data Eng.* **21**(9), 1263–1284 (2009)
21. Veale, M., Binns, R., Edwards, L.: Algorithms that remember: Model inversion attacks and data protection law. *Philos. Trans. R. Soc. A* **376**, 1–13 (2018)
22. Gama, J., Žliobaitė, I., Bifet, A., Pechenizkiy, M., Bouchachia, A.: A survey on concept drift adaptation. *ACM Comput. Surv.* **46**(4), 1–37 (2014)
23. Fuster, A., Goldsmith-Pinkham, P., Ramadorai, T., Walther, A.: Predictably unequal? The effects of machine learning on credit markets. *J. Finance* **77**(1), 5–47 (2022)
24. Thomas, L.C.: *Consumer Credit Models: Pricing, Profit and Portfolios.* Oxford University Press, Oxford (2009)
25. Pan, S.J., Yang, Q.: A survey on transfer learning. *IEEE Trans. Knowl. Data Eng.* **22**(10), 1345–1359 (2009)
26. Yang, Q., Liu, Y., Chen, T., Tong, Y.: Federated machine learning: Concept and applications. *ACM Trans. Intell. Syst. Technol.* **10**(2), 1–19 (2019)

27. Hardt, M., Price, E., Srebro, N.: Equality of opportunity in supervised learning. In: *Proc. Advances in Neural Information Processing Systems (NeurIPS)* (2016)
28. Óskarsdóttir, M., Bravo, C.: Multilayer network analysis for improved credit risk prediction. *Expert Syst. Appl.* **168**, 114–129 (2021)