A Lightweight Neural Network with Transformer to Predict Credit Default

1st Zongqi Hu
School of Computer Science and Engineering
Nanyang Technological University
Singpore
ZONGQI001@e.ntu.edu.sg

2nd Chai Kiat Yeo
School of Computer Science and Engineering
Nanyang Technological University
Singpore
ASCKYEO@ntu.edu.sg

***Abstract*—Financial institutions desperately need accurate credit default prediction tools to minimize losses, optimize lending and foster responsible borrowing. Existing methods often rely on complex ensemble approaches which are computationally intensive. This paper proposes a relatively lightweight solution: a simple encoder-only transformer which extracts latent features directly from credit card transaction data, followed by a simple neural network for the credit default prediction. The proposed approach achieves comparable performance on the American Express (AMEX) Default Prediction Kaggle Competition dataset against state-of-the-art (SOTA) ensemble techniques such as 3-Ensemble Modules with Gradient Boosting Decision Tree (GBDT) and 3-LightGBM with Dropouts meet Multiple Additive Regression Trees (DART) and pseudo-labelling ensemble methods. It achieves better performance than LightGBM based method.**

***Index Terms*—Credit default, Time series prediction, Machine learning, Transformer**

I. Introduction

Credit default risk refers to the probability of a borrower's failure to fulfill their debt obligations within the stipulated time-frame [1]. The prevalence of credit failures poses a significant threat to the financial health of lending institutions, potentially hindering the expansion and efficiency of the entire financial sector [2]. With accurate credit default prediction tools, financial institutions can minimize losses, optimize lending practices, and ultimately, promote responsible borrowing. This, in turn, fosters financial stability for both lenders and borrowers, leading to a healthier and more inclusive financial ecosystem for everyone [3].

While established models like SVM, LR and RF played a key role in borrower risk assessment [4], [5], recent SOTA research has shifted towards LightGBM, XGBoost and ensemble methods to extract deeper insights from financial data and enhance prediction accuracy [6]–[8]. These methods typically involve heavy computation given the various ensembles to push up the model performance. Using the AMEX dataset, Gan et al. proposed a prediction model built on lightGBM with some specific feature engineering [8]. Wang proposed a machine learning method based on Dropouts meet Multiple Additive Regression Trees (DART) and pseudo-labelling with multiple LightGBM ensemble [6]. Guo et al. proposed a multi-modal ensemble prediction approach comprising a LightGBM module, a XGBoost module using GBDT and a Local Ensemble module using lightGBM and CatBoost [7].

This paper proposes a lightweight transformer based neural network to extract temporal and other latent information from the benchmark AMEX customer data, which achieves comparable with if not better performance than the above-mentioned SOTA approaches.

II. Approach

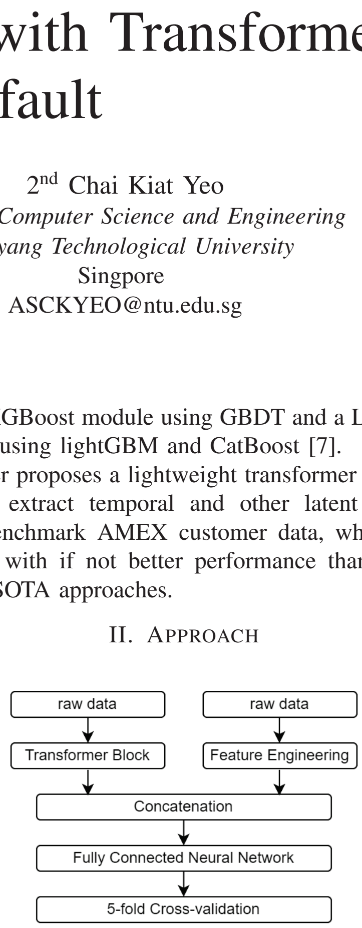

Fig. 1: Architecture overview

While existing studies heavily rely on tree-based models and ensemble approaches, their limitations in capturing temporal dependencies and extracting latent features from sequential data can impact their effectiveness in credit default prediction. This paper proposes a novel approach utilizing a transformer-based neural network that overcomes these limitations. Fig. 1 shows the architecture overview of the proposed model.

A. Transformer Block

The transformer block features two encoder layers with two attention heads specifically tailored for capturing the temporal dynamics and other latent features in the credit card data. Its self-attention mechanism effectively learns latent relationships between transactions, uncovering hidden patterns which can improve the prediction performance. As not every customer has the same length of series raw data, a special pooling mechanism is designed to take the encoder output corresponding to the last available series data for each customer.

B. Feature Engineering

Similar to the SOTA models, several feature engineering methods have been adopted to expand the data features to

979-8-3503-5409-6/24/$31.00 ©2024 IEEE
DOI 10.1109/CAI59869.2024.00015

capture more dynamic dependency between the multiple variables. One-hot encoding is performed for categorical features. Minimum, maximum, sum, difference, rank and standard deviation of the series raw data (such as spending amount) within a defined time windows (eg. 3 or 6 months) are calculated to capture the potential financial trends. These engineered features provide domain-specific insights leading to a more robust model. The engineered features are fed into a two-layer fully connected neural network with dropout mechanism to extract the hidden features if any.

C. Concatenation and Fully Connected Neural Network

The extracted features from both the transformer and the feature engineering are then concatenated and fed into a three-layer fully connected neural network for the final binary classification task. This combined approach leverages the strengths of both methods, allowing the network to learn complex relationships between the extracted features to predict the credit default risk with greater accuracy.

D. 5-fold Cross Validation

To ensure a robust evaluation, a 5-fold cross-validation strategy is implemented. This iterative process repeatedly trains and tests the model on different splits of the data, providing more reliable and generalizable performance estimates.

III. Experiments and Results

A. Dataset

As financial data are not readily available for research purpose, we leverage on the AMEX Credit Default Kaggle Competition dataset (released in 2022 May) for credit card default prediction [9]. There are 458,913 and 924,621 customer records available in the training and datasets respectively. The test dataset is further split into 51% public test set for the competition paricipants to verify their models and 49% private test set as a final evaluation set for the competition ranking. Each customer record has over 200 rich and multi-temporal features across five categories like Delinquency, Spend, Balance etc. with a default label that indicates whether the customer will default after a 18-month observation window.

B. Evaluation Metric

The Kaggle competition's evaluation metric, denoted by M, is adopted to facilitate direct comparison with existing research. This composite metric captures the overall predictive performance by averaging two sub-measures: the Normalized Gini Coefficient (G) and the default rate captured at a threshold of 4% (D).

$$M = 0.5(G + D)$$

The default rate recorded at 4% corresponds to the percentage of defaults identified within the top 4% of the predictions. To adjust for down-sampling, the negative labels of both sub-metrics are given a weight of 20.

C. Experimental Results

The proposed and benchmark models are evaluated on both the public and private test datasets and the results are summarised in Table I. As an ablation study, the proposed model with transformer only and the proposed model with feature engineering (FE) only are also evaluated.

TABLE I: Results comparison

Model	Public Test Set	Private Test Set
GBDT Ensemble ([7])	0.8013	0.8087
DART Ensemble ([6])	0.8003	0.8077
LightGBM ([8])	-	0.8010
Transformer based Neural Network (Ours)	0.7969	0.8047
Transformer Only (Ours)	0.7890	0.7989
FE Only (Ours)	0.7946	0.8035

The proposed model achieves comparable performance as the SOTA GBDT and DART ensemble methods while outperforming a LightGBM method despite its lightweight design. This demonstrates the capability of the transformer in extracting the latent features from the time series dataset even though the time samples of the dataset are very short. The ablation study shows the importance of augmented features in boosting the performance of the transformer only structure and vice versa, underscoring the strength of their combination.

IV. Conclusion

In this paper, we propose a lightweight transformer based neural network to extract temporal and latent features for credit default prediction from a dataset with very short time samples. Experiments performed on the benchmark AMEX dataset show that the proposed model is able to achieve comparable results to the SOTA heavy ensemble solutions and outperform LightGBM method. Future study could focus on optimising the transformer design to further improve its performance on short sequence dataset.

References

[1] H. Kim, H. Cho, and D. Ryu, "An empirical study on credit card loan delinquency," Economic systems, vol. 42, no. 3, pp. 437–449, 2018.

[2] M. Leo, S. Sharma, and K. Maddulety, "Machine learning in banking risk management: A literature review," Risks, vol. 7, no. 1, p. 29, 2019.

[3] X. Liu, Y. Li, C. Jiang, Z. Wang, F. Zhao, and J. Wang, "Attentive feature fusion for credit default prediction," in 2022 IEEE 25th International Conference on Computer Supported Cooperative Work in Design (CSCWD). IEEE, 2022, pp. 816–821.

[4] S. Menard, Applied logistic regression analysis. Sage, 2002, no. 106.

[5] L. Zhu, D. Qiu, D. Ergu, C. Ying, and K. Liu, "A study on predicting loan default based on the random forest algorithm," Procedia Computer Science, vol. 162, pp. 503–513, 2019.

[6] Haoming Wang. 2023. Forecasting Credit Card Defaults Using Light Gradient Boosting Machine with Dart Algorithm. In Proceedings of the 2022 5th Artificial Intelligence and Cloud Computing Conference (AICCC '22). Association for Computing Machinery, New York, NY, USA, 207–212.

[7] K. Guo et al., "Credit Default Prediction on Time-Series Behavioral Data Using Ensemble Models," 2023 International Joint Conference on Neural Networks (IJCNN), Gold Coast, Australia, 2023, pp. 01-09

[8] Zhiren Gan, Junyuan Qiu, Fuli Li, and Qian Liang. A light-gbm based default prediction method for american express. EAI, 8 2023.

[9] Addison Howard, AritraAmex, Di Xu, Hossein Vashani, inversion, Negin, Sohier Dane. (2022). American Express - Default Prediction. Kaggle. https://kaggle.com/competitions/amex-default-prediction