ISBN: 978-981-94-1156-6
2024 the 14th International Workshop on Computer Science and Engineering (WCSE 2024)
doi: 10.18178/wcse.2024.06.027

Developing a Credit Scoring Model for Evaluating Credit Risk in the Nepali Setting

Suresh Gautam+ and Bhoj Raj Ghimire

Nepal Open University, Nepal

Abstract. Access to credit is vital for economic development, yet Nepal lacks tailored credit scoring models, particularly for retail and micro-lending, due to limited credit history. In this study, we address this gap by developing a locally relevant credit scoring model using data from Foneloan, a digital lending platform. Leveraging machine learning techniques and Python libraries, we identify influential features, determine creditworthiness parameters, and select appropriate algorithms (ADA, DT, ET, RF, XGBoost). Our analysis reveals a significant decrease in bad rate or non-performing loan (NPL) across various score ranges with modified credit scoring model compared to the standard globally used model indicating enhanced loan performance and creditworthiness. This suggests that borrowers with higher credit scores are less likely to default, reflecting better credit management practices. Overall, our study contributes to improving credit risk management in Nepal, fostering economic development and financial inclusion.

Keywords: Credit Risk Management, Credit Scoring, Digital Lending

1. Introduction

Nepal has witnessed significant advancements in its financial sector in recent years, leading to an increasing number of financial institutions offering credit products to consumers and businesses. However, these advancements have brought challenges related to credit risk management, including high default rates and inadequate risk assessment methodologies. Current credit scoring models utilized in Nepal predominantly depend on pre-existing models that have been successful in other markets. However, there has been limited exploration into optimizing these models for the local context or integrating additional data points that could enhance their effectiveness.

1.1. Problem Statement

The lack of a comprehensive credit scoring framework tailored to the Nepalese context has implications for both lenders and borrowers. Financial institutions may face increased risk exposure due to inadequate risk assessment practices, leading to higher provisioning requirements and potential financial losses. On the other hand, borrowers may encounter difficulties in accessing credit or may be subjected to unfair lending practices due to the absence of a standardized and transparent credit evaluation process. In the Nepalese context, the setup of credit scoring is overseen by Nepal's first and only private credit bureau, Karja Suchana Kendra Limited (KSKL). KSKL has implemented a global standard credit scoring framework since 2012 [1]. There is a notable absence of data-driven disclosure mechanisms to identify the most influential features, predictive characteristics, and their respective weightage in credit score modeling in the Nepalese context. This deficiency hampers the development of accurate and reliable credit scoring models tailored as per the local market dynamics and optimize efficiency to gatekeeping the overdue or defaults [2].

Among the 30.896 million population in the country by end of 2023 [3], there are merely 1.867 million loan accounts across financial institutions of Class A, B, and C, and 2.735 million bowers at microfinance institutions in Nepal [4]. Furthermore, only the lending footprint population would have access to credit reports and credit scores available at the credit bureau, highlighting the limited reach of credit scores in the nation.

Credit Scoring framework enables compliance with regulatory requirements, which helps maintain trust and confidence among stakeholders, including investors and customers. The absence of automatic credit risk

+ Corresponding author.
E-mail address: sureshpro@gmail.com

assessment and the creation of credit files for lending purposes pose significant challenges for financial institutions in Nepal. Without automated systems in place, assessing the creditworthiness of borrowers becomes labour-intensive and prone to errors[5]. Additionally, the lack of comprehensive credit files hinders the institution's ability to make informed lending decisions and effectively manage credit risk exposure.

1.2. Objectives

The primary objective of this study is to develop a robust and locally relevant credit scoring model for the Nepalese context. Specifically, the objectives include identifying relevant features for credit risk assessment, determining the most influential parameters for creditworthiness assessment in the Nepalese retail lending market, developing a credit score framework, and identifying suitable machine learning algorithms for loan classification.

2. Related Work

2.1. Credit Scoring Model

The idea of credit scoring was initially introduced by the Fair Isaac Corporation (now known as FICO) in 1956 [6]. Subsequently, credit scoring services have been provided by various credit bureaus, including FICO, Experian, and TransUnion, among others and FICO scores are used by 90% of top lenders[7]. According to [8] [9] [10], credit scoring model employs scorecard method to evaluate the data and generating the score, FICO 8 score ranges between 300 and 850 and it's calculated by five key metrics, 35% payment histories, 30% amount owed, 15% length of credit history, 10% new credit, and 10% credit mix.

As credit score modelling evolves and becomes increasingly integral to credit risk management, novel approaches utilizing various machine learning models, including different classifiers and clustering techniques w/o fuzzy assignment are being proposed in [11] [12] [13], Base classifiers KNN, LR, RF, GBDT, LDA, SVM, NN, AdaBoost and DT are applied to the credit scoring problem. Ensemble classifiers are constructed from the base classifier, including combination of multiple classifiers clustering procedure with fuzzy assignment added to the ensemble classifier to test effectiveness of clustering CF-Ens. Genetic algorithm (GA) is also added to the CF-Ens model to form CF-GA-Ens model, to test the effectiveness of using GA to dynamically select classifiers in [11]. [13] proposed a combination of natural language explanations and visualizations to enhance the explainability of the machine learning granting scoring model. In [11] [12] [13], performance of the model was evaluated using cross-fold validation and metrics such as accuracy, precision, specificity, F1 score, and area under the curve (AUC). Furthermore in [13], model adjustment also involves fitting a logistic regression (LR) model with a standard set and a balanced set, hyperparameter optimization using a genetic algorithm, maximizing balanced accuracy (BAC), and applying K-fold cross-validation with same machine learning algorithms and added BAC for model selection, Kolmogorov-Smirnov (KS) statistic).

2.2. Feature Extraction and Prediction Modeling

[14] [15], followed a comprehensive approach to feature engineering and classification by leveraging ensemble and ensemble combined with feature selection to enhance predictive performance in credit risk assessment and for predicting default risk. [14] applied multi-faceted approach that integrates tree-based methods, deep learning techniques, and ensemble learning. [15], conducted a feature extraction process by merging information value (IV) and the spearman-Boruta algorithm, leading to the establishment of a robust credit evaluation index system rooted in the 5C (i.e. character, capacity, capital, collateral, and condition) credit evaluation principle. The Stacking algorithm emerged as a pivotal component, enabling the integration of diverse base classifiers like Artificial Neural Networks (ANN), Random Forest (RF), Adaptive Boosting (AdaBoost), and Gradient Boosting (XGBoost). This ensemble approach effectively discriminates loan defaults in network lending scenarios, leveraging the collective strengths of individual classifiers to deliver reliable predictions. In [16], statistical models such as logistic regression (LR), Bayesian classifier, and linear discriminant analysis were employed and DT, RF, LGBM, ANN, and CNN for data analysis and AI modelling respectively. [17], utilized feature selection through restricted Boltzmann Machines (RBM) combined with deep learning Methodology.

2.3. Methodological Approach

It's experimental research that aims to develop a robust credit scoring model tailored to the Nepalese retail lending market, with the objective of evaluating credit risk effectively. Experimental design is chosen as the customization and evaluation of a credit score model specific to the Nepalese context have not been undertaken so far and quantitative approach is employed as it allows greater explainability. Overall system architecture and context level diagram shown in the Figure 1 below.

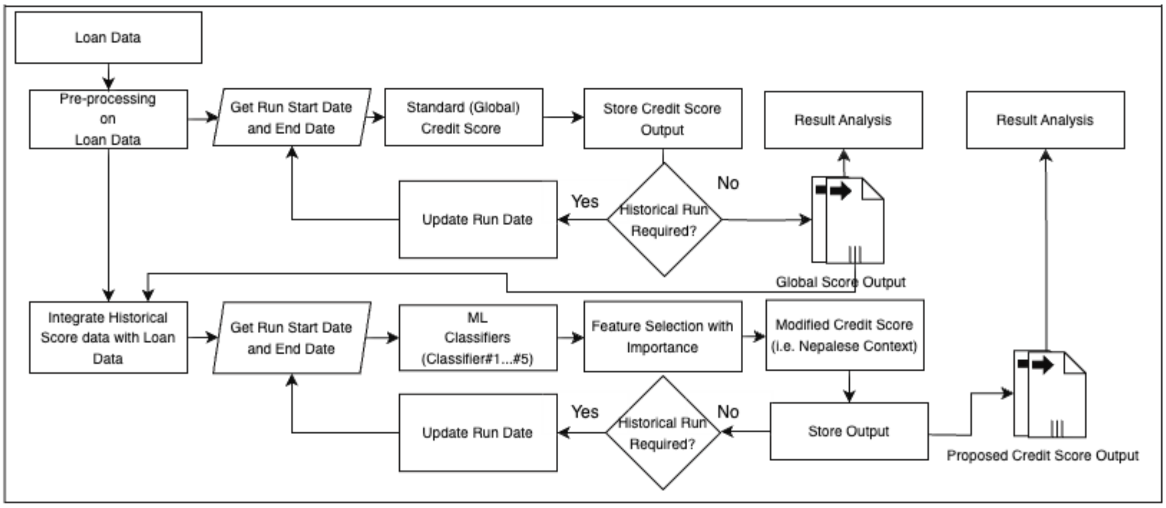

Fig. 1: Context Level Diagram of Proposed Solution

2.4. Data Pre-processing

One-hot encoding data transformation is used in converting categorical feature and feature scaling to scale into [0,1] or the similar peer variables range. Filling in missing values with estimated values based on mean and mode imputation are applied to numerical and categorical data respectively. SMOTE (Synthetic Minority Over-sampling Technique) is employed to mitigate class imbalance by creating synthetic samples for the minority class.

2.5. Toolset and Environment

The processing environment for this study primarily relies on Python scripts and libraries, leveraging key libraries such as pandas, numpy, pycarret, sklearn, pandas profiling, jupyter notebook, and jupyter notebook toolsets are predominantly utilized because of their open-source nature, extensive community support, well-documented resources, and a vast array of available libraries and their regular updates.

2.6. Machine Learning Classifiers

Tree-based classifiers ADA boosting classifier, gradient boosting classifier, extreme gradient boosting classifier, decision tree classifier, and random classifier well-suited for identifying feature importance in datasets, providing valuable insights and provide transparency to access features and their importance. Also, these classifiers are used in [11] [12] [13] [14] [15] for feature selection and prediction modelling.

2.7. Feature Selection

Feature selection is conducted using the feature importance determined by individual classifiers to identify the most relevant features from the existing pool. We use the average rank aggregation method to create a unified ranking of feature importance when we have multiple sets of feature importance from different classifiers, as described in [18]. Furthermore, features with an importance greater than 0.01 (equivalent to 1%) are considered during the feature selection process from the unified feature importance ranking. This selection encompasses features that contribute to 94% of the total importance, with the remaining 6% being incorporated as constraints in the regression equation.

2.8. Cross-fold generation and Evaluation

Cross-fold generation, also known as cross-validation, is a technique used to assess the performance of a predictive model. In the evaluation process, metrics such as accuracy, AUC, recall, precision, and F1 score are utilized to assess the performance of the classifier model in the process of feature selection and Mann-Kendall test to assess trends of bad rate or NPL in both standard and proposed modified credit score implementation.

2.9. Dataset Overview

Utilized a masked dataset sourced from Foneloan, a prominent digital lending platform operating in Nepal. This dataset comprises a comprehensive footprint of digital lending activities, offering valuable insights into various aspects of loan processing and management. It consists of 22 variables and 157,999 observations. The loan processed date spans from November 29, 2021, to February 10, 2024, providing a rich temporal perspective. Additionally, dataset having majority of loans are 128,330 settled, followed by 2,559 active, 2,983 overdue, and 1096 default statuses. Moreover, dataset comprises two primary loan types: 1 month loan and 12 months EMI loan, with 53,394 and 104,605 respectively. Furthermore, loan tenures vary, ranging from 1 month to 12 months.

3. Results and Discussion

In this section, describe how the results of determining the most influential parameters, standard and modified credit score implementation along their performance, and discussion about the results.

3.1. Determination of the Most Influential Parameters

Figure 2 shows the feature importance scores generated by combining the rankings from all classifiers using the average rank aggregation method. Higher scores indicate greater importance of the feature in predicting the target variable and lower importance score indicates lower importance. Features with an importance greater than 0.01 (equivalent to 1%) are considered during the feature selection process from the unified feature importance ranking, it contributes to 94% of total importance with the remaining 6% being incorporated as a constraint in the regression equation.

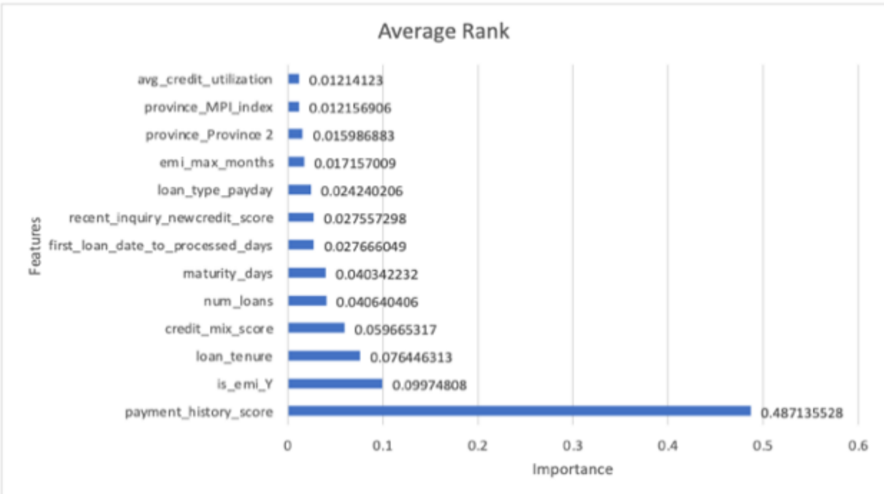

Fig 2: Relevant Features with Importance

3.2. Comparative Analysis

Comparison of standard credit score with domestic and global bad rates:

In Figure 3, the first line includes a table that provides insights into the credit score implemented with the same standard configuration. The table includes the credit score range in the first column of the domestic dataset, metrics related to non-performing loans (NPLs) with domestic bad rates in the second column, global actual bad rates in the third column, and global expected bad rates in the fourth column. In a subsequent line in Figure 3, the second graph presents the distribution from the first table in a graphical form. With the NPL rate depicted in y-axis, the figure indicates a negative correlation between credit score range and NPL, suggesting that as the credit score range increases, the NPL decreases. Second graph in Figure 3 shows Mann-Kendall test yields a Tau value of -0.58, it suggests a moderate negative correlation or trend within the data distribution. This implies a discernible downward tendency in the bad rate or NPL percentages across various score ranges.

Credit Score Range (Stanard)	Domestic Bad Rate (Above 90Days) % [NPL/Active Loan]	Global Actual Bad (NPL) %	Global Expected Bad (NPL) Rate
300-621.8	2.32%		
621.8-636.2	1.63%		
636.2-654.2	3.27%		
654.2-672.2	0.00%	26.30	18.00
672.2-690.2	0.00%	15.30	14.00
690.2-708.2	0.00%	6.20	10.00
708.2-726.2		6.50	8.00
726.2-744.2		4.20	5.00
744.2-762.2		3.80	4.00
762.2-780.2		1.90	2.00
780.2-798.2		1.10	0.80
798.2-850		0.00	0.50

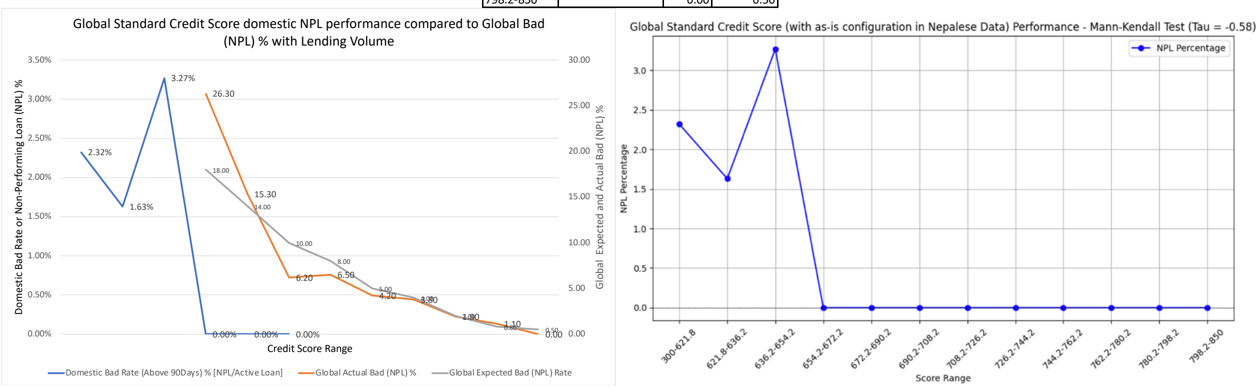

Fig. 3: Global Standard Credit Score Bad Rate in Domestic Dataset: Table on the first line, and Graph and Mann-Kendall Test Results on the second line.

Comparison of modified credit score with domestic and global bad rates:

In Figure 4, the first line includes a table that provides insights into the credit score implemented with the same standard configuration taking the features from section 3.1 in domestic dataset. The table includes the credit score range in the first column, metrics related to non-performing loans (NPLs) with domestic bad rates in the second column, global actual bad rates in the third column, and global expected bad rates in the fourth column. In a subsequent line in Figure 4, the second graph presents the same distribution from the table in a graphical form. With the NPL rate depicted in y-axis, the figure suggests a negative correlation between credit score range and NPL, indicating that as the credit score range increases, the NPL decreases. Second graph in Figure 4 shows Mann-Kendall test score of -0.75, we observe a strong negative correlation or trend in the data distribution, we observe a strong negative correlation or trend in the data distribution.

Credit Score Range (Modified)	Domestic Bad Rate (Above 90Days) % [NPL/Active Loan]	Global Actual Bad (NPL) %	Global Expected Bad (NPL) Rate
300-621.8	2.34%		
621.8-636.2	2.31%		
636.2-654.2	3.36%		
654.2-672.2	2.31%	26.3	18
672.2-690.2	1.98%	15.3	14
690.2-708.2	2.98%	6.2	10
708.2-726.2	1.80%	6.5	8
726.2-744.2	1.22%	4.2	5
744.2-762.2	1.58%	3.8	4
762.2-780.2	0.00%	1.9	2
780.2-798.2	0.00%	1.1	0.8
798.2-850	0.00%	0	0.5

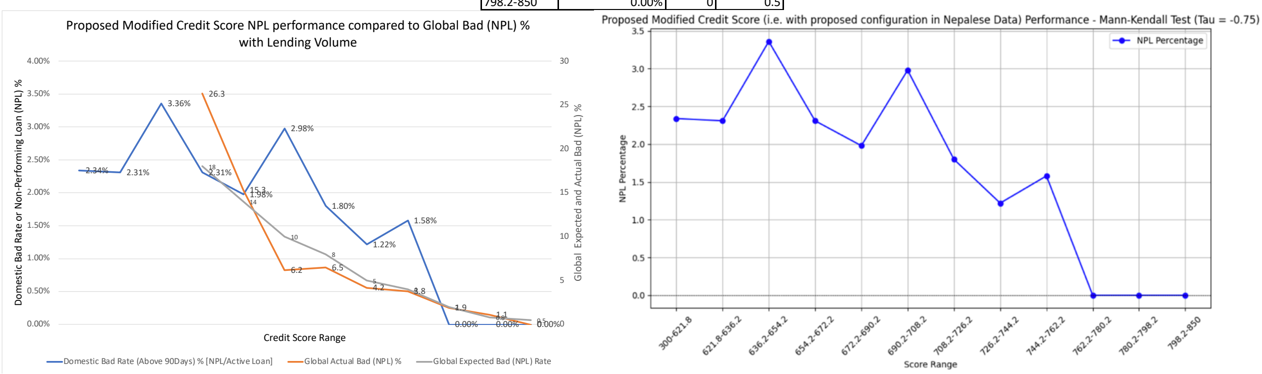

Fig. 4: Proposed Modified Credit Score Bad Rate in Domestic Dataset: Table on the first line, and Graph and Mann-Kendall Test Results on the second line.

Above comparison shows modified credit perform well with -0.75 Mann-Kendall test score over standard credit score with -0.58 Mann-Kendall test score in domestic Foneloan dataset, it suggests a significant downward trend in the Bad Rate or NPL (Non-Performing Loan) percentages across various score ranges in modified credit score. The implication is that as the score ranges increase, there is a batter tendency for the bad rate or NPL percentages to decrease, indicating a potential enhancement in loan performance or creditworthiness. It implies that borrowers with higher credit scores in modified credit score are less likely to default on their loans, possibly reflecting better credit management practices or more precise risk assessment methodologies.

3.3. Recommendation

Credit bureaus should adjust credit scoring systems to include localized behaviours and establish flexible frameworks for ongoing analysis and model adjustments. This study's findings emphasize the necessity of this process. Also, regulatory and other relevant stakeholders should take measures should expand beyond retail loan and credit card data to include all financial commitments like utility bills, micro-loans, and credit union memberships, establish a policy that promotes the development of credit histories from an early age by encouraging individuals to maintain discipline in repaying any financial obligations and commitments, promote automated credit risk assessment and unbiased lending decisions, with scalability and refinement based on results, encouraging individuals to maintain disciplined financial habits from a young age is vital for building comprehensive credit histories.

4. Conclusion

In conclusion, this study addresses the critical need for a locally relevant credit scoring model tailored to the Nepalese context. By leveraging a dataset from Foneloan and employing advanced machine learning techniques, we have identified influential features, determined creditworthiness parameters, and built a modified credit score, and then run with the historical data. Our analysis on the results of modified credit score reveals a significant downward trend in non-performing loan (NPL) percentages, across various credit score ranges, it achieved Mann-Kendall test score of -0.75 as compared to Mann-Kendall test score of -0.58 with standard credit score model run in the same dataset, indicating potential enhancements in loan performance and creditworthiness. This suggests that borrowers with higher credit scores are less likely to default on their loans, reflecting better credit management practices or more precise risk assessment methodologies. Overall, our study contributes to improving credit risk management in Nepal's financial sector, facilitating economic development and financial inclusion. Further research and implementation of the proposed credit scoring model can help address the challenges and opportunities unique to the Nepalese lending landscape, ultimately benefiting both lenders and borrowers.

5. References

[1] "Karja Suchana Kendra Limited Annual Report 2013." Karja Suchana Kendra, Credit Information Bureau Nepal, 2013. Accessed: May 05, 2024. [Online]. Available: https://nepalindata.com/media/resources/items/17/bAnnual_report_2013.PDF

[2] X. Bai and Z. Zhao, "An Optimal Credit Scoring Model Based on the Maximum Default Identification Ability for Chinese Small Business," *Discrete Dyn. Nat. Soc.*, vol. 2022, pp. 1–14, Jan. 2022, doi: 10.1155/2022/1551937.

[3] WorldMeter, "Nepal Population." [Online]. Available: https://www.worldometers.info/world-population/nepal-population/

[4] Nepal Rastra Bank, a Central Bank of Nepal, "Banking and Financial Statistics." [Online]. Available: https://www.nrb.org.np/contents/uploads/2024/02/Poush_2080_Publish-2.xlsx

[5] N. Biswas, A. S. Mondal, A. Kusumastuti, S. Saha, and K. C. Mondal, "Automated credit assessment framework using ETL process and machine learning," *Innov. Syst. Softw. Eng.*, Dec. 2022, doi: 10.1007/s11334-022-00522-x.

[6] M. A. Poon, "What lenders see–: a history of the Fair Isaac scorecard," PhD Thesis, UC San Diego, 2012.

[7] "FICO Score Versions," FICO Score Versions.

[8] M. Fowlie, "FICO Reverse Engineering." Accessed: May 04, 2024. [Online]. Available:

https://medium.com/@mfow/reverse-engineering-fico-8-d2d68315d20

[9] FICO, "Credit Scoring Primer." Accessed: May 04, 2024. [Online]. Available: https://ficoforums.myfico.com/t5/Understanding-FICO-Scoring/Credit-Scoring-Primer-pub-5-17-20/m-p/6023348/highlight/true

[10] N. Siddiki, "Intelligent Credit Scoring," *SAS Inst.*, 2017.

[11] H. Zhang, H. He, and W. Zhang, "Classifier selection and clustering with fuzzy assignment in ensemble model for credit scoring," *Neurocomputing*, vol. 316, pp. 210–221, 2018.

[12] X. Chen, S. Li, X. Xu, F. Meng, and W. Cao, "A Novel GSCI-based Ensemble Approach for Credit Scoring," *IEEE Access*, 2020.

[13] M. J. Ariza-Garzon, J. Arroyo, A. Caparrini, and M.-J. Segovia-Vargas, "Explainability of a Machine Learning Granting Scoring Model in Peer-to-Peer Lending," *IEEE Access*, 2020.

[14] H. He and Y. Fan, "A novel hybrid ensemble model based on tree-based method and deep learning method for default prediction," *Expert Syst. Appl.*, vol. 176, p. 114899, 2021.

[15] Z. Kun, F. Weibing, and W. Jianlin, "Default identification of p2p lending based on stacking ensemble learning," in *2020 2nd International Conference on Economic Management and Model Engineering (ICEMME)*, IEEE, 2020, pp. 992–1006.

[16] P.-C. Ko, P.-C. Lin, H.-T. Do, and Y.-F. Huang, "P2P lending default prediction based on AI and statistical models," *Entropy*, vol. 24, no. 6, p. 801, 2022.

[17] V.-S. Ha, D.-N. Lu, G. S. Choi, H.-N. Nguyen, and B. Yoon, "Improving credit risk prediction in online peer-to-peer (P2P) lending using feature selection with deep learning," in *2019 21st International Conference on Advanced Communication Technology (ICACT)*, IEEE, 2019, pp. 511–515.

[18] S. Najdi, A. A. Gharbali, and J. M. Fonseca, "Feature ranking and rank aggregation for automatic sleep stage classification: a comparative study," *Biomed. Eng. OnLine*, vol. 16, no. S1, p. 78, Aug. 2017, doi: 10.1186/s12938-017-0358-3.