Stacking Ensemble LearninG Approach for Credit Rating of Bank Customers

Qinyu Guo
College of Computer and Information Science & College of Software, Southwest University, Chongqing, China

Keywords: Bank Credit Scores, Machine Learning, Stacking Model.

Abstract: The banking industry has experienced tremendous growth and change in recent years, creating new challenges and opportunities for credit assessment and management. In this context, accurately and efficiently assessing customer credit risks has become the key to the success of the banking business. A financial risk approval model based on stacking technology is proposed in response to this demand. The model starts by selecting a data set containing multiple bank user features. After a series of steps, such as data preprocessing, feature selection, preliminary model training, and model optimization, it finally forms a credit assessment model with high prediction accuracy. During the model training process, various machine learning algorithms were used for comparison, including neural networks, random forests, decision trees, naive Bayes, etc., and the algorithms were improved through stacking technology to achieve higher accuracy and Area Under Curve (AUC). In addition, based on the stacked model's prediction results, each customer's credit score is also calculated, and the distribution of customers with different credit score segments is displayed through visualization technology. This provides financial institutions with detailed information about their customers' credit risks, helping them formulate more reasonable lending policies and interest rates. Experimental results show that compared with other models, the proposed superposition-based risk approval model improves the joint loan approval rate by about 6% on the actual data set, proving its effectiveness and feasibility in financial risk assessment.

1 INTRODUCTION

Assessing the risks of lending money to a person or a business is the credit rating process. One of the main methods financial institutions use to evaluate their operational risks is credit rating, which tries to detect applicants with poor credit who may have a high likelihood of defaulting (Jiang and Packer 2019).

With the increasing market uncertainty, investors face more significant risks when investing large amounts. It has become an urgent issue to accurately assess and predict the risks and returns of credit products. Financial institutions have vast data about borrowers, including their historical borrowing and repayment records, economic status, social media activity, and consumption behavior. However, with the rapid growth of credit-related financial product markets, the challenge is accurately extracting valuable information from this complex and diverse data (Musdholifah et al 2020). Certain data features' correlation with credit assessments may need to be more stable and can even change over time.

In consideration of this, banks play an essential role in determining credit risk. Before approving a loan, evaluating a borrower's credit history is necessary to identify potential high and low risks (Kadam et al 2021). Then, machine learning (ML) algorithms, which enable systems to decipher patterns and make data-driven predictions autonomously have emerged as a promising solution for assessing the likelihood of loan defaults (Musdholifah et al 2020). However, to further accurately predict bank user loan risks, the model needs further enhanced (Beutel et al 2019).

Therefore, the Ensemble learning method is adopted, mainly including Bagging (Bootstrap Aggregating), creating multiple sub-datasets through random sampling, and training each subset independently (Uddin et al). The final result is based on each learner's average or majority vote (Erdal and Karahanoğlu 2016). Boosting: by introducing the learner step by step and adjusting based on the error of the previous step, the goal is to reduce the overall error and enhance the model's performance (Carmona et al 2019).

Ensemble learning methods offer a superior alternative to traditional machine learning techniques. They frequently achieve improved accuracy and lower

Guo, Q.
Stacking Ensemble Learning Approach for Credit Rating of Bank Customers.
DOI: 10.5220/0012801200003885

In *Proceedings of the 1st International Conference on Data Analysis and Machine Learning (DAML 2023)*, pages 274-278
ISBN: 978-989-758-705-4

the danger of overfitting by integrating the predictions of various models. Their intrinsic versatility enables the incorporation of multiple methods, improving generalization to unseen data and performance in high-dimensional settings (Uddin et al). Subsequently, the efficacy of these models is rigorously assessed using metrics such as accuracy and AUC, ensuring a comprehensive evaluation of their predictive capabilities (Wang et al 2023).

As a result, this research aims to combine the ensemble learning method and big data analysis to develop a new, accurate, and reliable credit assessment model. This model will help financial institutions and investors better understand investment opportunities and achieve a balance between risk and returns in a complex market environment.

2 RELATED WORK

Combining various model predictions, traditional Ensemble learning offers a distinct advantage over conventional techniques by enhancing accuracy and preventing overfitting. Methods like Voting, Bagging, and Boosting are pivotal in this domain.

Erdal and Karahanoğlu explored the determinants of profits for Development and Investment Banks in Turkey using bagging ensemble models (Erdal and Karahanoğlu 2016). Leveraging three tree-based machine learning models as base learners, their study revealed that ensemble models, specifically Bag-DStump, Bag-RTree, and Bag-REPTree, outperformed individual models in predicting the profitability determinants of Turkish banks.

Uddin et al. introduce a machine learning based loan prediction system better to identify qualified bank loan applicants (Uddin et al). Nine machine learning algorithms and three deep learning models were used, achieving enhanced performance with ensemble voting model techniques.

Carmona, Climent, and Momparler employ extreme gradient boosting techniques to predict bank failures in the U.S. banking industry systematically. It effectively identifies critical indicators related to bank defaults, which is crucial in determining a bank's vulnerability (Carmona et al 2019).

This study aims to exploit the complex hybrid capabilities of Stacking. Unlike other methods, this method utilizes multiple models to generate meta-features fed into another model for final prediction. This approach not only consolidates the strengths of each model but also compensates for their respective weaknesses, paving the way for more accurate and detailed credit assessment models.

Therefore, the primary data sources were obtained from the UCI Machine Learning Repository, which describes the dataset's attributes (South German Credit 2020). It highlights the effectiveness of ensemble learning, focusing on Stacking methods. A robust and comprehensive predictive model has been implemented using Random Forest and Gradient Boosting as the base models and Logistic Regression as the meta-model. A distinctive feature of the approach is to convert model-predicted probabilities into standardized credit scores, providing institutions with intuitive indicators. By combining meticulous data preprocessing, feature selection, and innovative applications of stacking and mapping techniques, this work offers a unique solution to the credit assessment field.

3 METHODOLOGY

This section delves into the study of bank credit prediction models and evaluates them. Then, the predictive model method was optimized and improved through ensemble learning. It provides a comprehensive breakdown of evaluation indicators, comparative analysis, and final credit scores.

3.1 Data Preprocessing

The data is looked up to identify potential missing values in the variables in the dataset. Fortunately, the data set is complete without missing values, so no additional data imputation step is required.

To ensure that the model remains unbiased, the Z-score method can be implemented to detect and address any outliers within the data. This method calculates the difference between an observation and the mean as several standard deviations. By setting a threshold of 3 standard deviations, any data points with a Z-score value more significant than three can be identified as outliers, which can then be safely removed.

3.2 Exploratory Data Analysis

Thermal correlation plots can be used to identify relationships between features in data. Analyzing these plots makes it possible to discover correlations between different variables, making them a valuable tool for data analysis. Through Fig. 1, it is confirmed that the relationship between these features is weak, so multicollinearity problems will not occur in subsequent model training.

3.3 Feature Selection

To ensure that all features are on the same scale, the data was standardized using StandardScaler. This ensures the model has even weights across parts and helps speed up the training process.

Feature selection was then performed, and the importance of all features was evaluated using a random forest classifier. Based on the feature importance results, only the top 10 most important components are retained to reduce the complexity of the model and avoid overfitting. These ten features (Fig. 2) are believed to predict a customer's credit risk better. The filtered critical feature data is saved to a new file for subsequent use.

3.4 PreliminarY Model Training

The research embraced four distinct supervised learning algorithms to devise an accurate credit scoring model to create a precise credit scoring model (Table 1). These encompassed Neural Networks, known for their prowess in recognizing intricate patterns; Random Forests, revered for their ensemble-based approach; Decision Trees, celebrated for their transparent decision-making structure; and Naive Bayes, distinguished for its probabilistic foundations. Each of these algorithms underwent meticulous parameter tuning to hone their performance. Their effectiveness was evaluated using two pivotal metrics: the Area Under the Curve (AUC) and accuracy. While the AUC offered insights into the model's overall performance across varied classification thresholds, accuracy provided a snapshot of the model's success rate in making correct predictions.

3.5 Model Improvement: Stacking Model

Stacking technology is employed to improve the model's predictive accuracy, a form of ensemble learning combining multiple machine learning algorithms to achieve better predictive performance. This study used random forest and gradient boosting as basic models. These base models are independently trained on the data and make individual predictions. However, no simple majority vote is taken, or these predictions are averaged.

Instead, the predictions from the random forest and gradient boosting models are used as new “meta-features” as input to subsequent logistic regression meta-models. Essentially, this meta-model is trained to make final predictions based on the predictions of the base model. This hierarchical arrangement of models effectively captures the respective strengths of the base models while compensating for their weaknesses.

3.6 Credit Score Calculation

The conversion of model-predicted probabilities into actual credit scores is achieved by applying the following mapping method. This method utilizes the prediction results obtained from the aforementioned stacked model for each individual customer.

$$C = A - B \times S(P), \quad S(P) = \frac{P - \min(P)}{\max(P) - \min(P)} \qquad (1)$$

A represents a fixed constant set at 300, which serves as the fundamental baseline for the credit score. B, on the other hand, is a constant with a value of 500, usually adjusted by logarithmic properties to accommodate diverse data distributions and specific requirements. P signifies the probability of high credit risk computed through the stacking process. This formula facilitates the mapping of potential outcomes to a credit score range spanning from 300 to 800. This, in turn, furnishes financial institutions with a more intricate and nuanced customer credit rating assessment.

3.7 Visualization of Results

A bar chart serves as a valuable visual tool, employed to effectively illustrate the distribution of customers across a spectrum of diverse credit scores. This graphical representation imparts a more comprehensive understanding of the model's output, shedding light on the precise count of customers within each distinct scoring segment and, in particular, how it delineates between the issuance of good and bad credit within each of these individual segments.

4 EPECTIONTAL RESULT

Within this section, the model's performance is subjected to a comprehensive comparative evaluation, utilizing the dataset. Furthermore, the conclusive bank credit scoring results are delineated and presented.

4.1 Dataset

The dataset chosen for this research study offers a comprehensive insight into the credit risk associated with bank customers. It comprises a total of 1,000 entries, consisting of 700 instances characterized by good credit and 300 instances associated with bad credit. This dataset encompasses 20 predictor variables that encompass a wide range of financial, professional, and personal background information about the customers under analysis.

4.2 Result

As depicted in Fig. 1, the correlation diagram reveals that the majority of variables exhibit weak correlations, as indicated by the coloration being close to neutral. This suggests that these variables maintain a relatively high degree of independence from each other. This independence is advantageous for model accuracy, as highly correlated variables can give rise to multicollinearity problems. Notably, the diagram does not display any dark red or dark blue blocks, signifying the absence of strong correlations between variables.

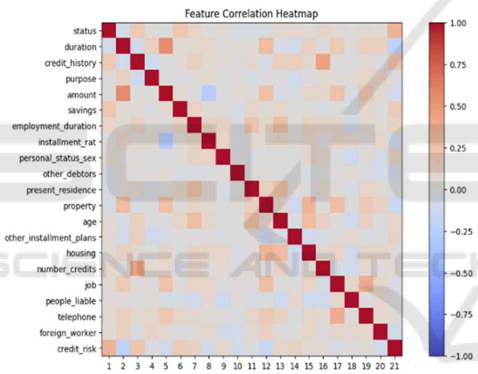

Figure 1: Correlation Heatmap (Picture credit: Original).

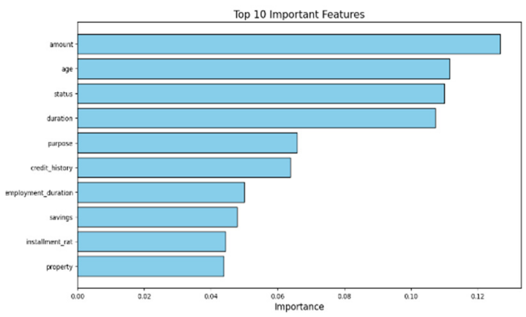

Figure 2: Top 10 Important Features (Picture credit: Original).

As shown in Fig. 2, outliers are successfully handled, and the data is normalized through data preprocessing. This ensures that the model has even weights on all features. Furthermore, during the feature selection process, we used the feature importance of the random forest classifier to retain only the most essential ten elements, reducing the model's complexity and effectively avoiding overfitting.

Model training: Training of the preliminary model provided good results in terms of AUC and accuracy. However, employ stacking techniques to improve the accuracy of predictions further. They combine random forest and gradient boosting as the base model and adopt logistic regression as the meta-model to enhance overall performance (Table 1).

Table 1: Performance of the various algorithms.

Algorithm	Accuracy	AUC
Model Stacking	93.71%	88.33%
Decision Tree	73.45%	70.03%
Random Forest	77.56%	75.45%
Logistic Regression	79.43%	78.74%
Neural Networks	73.45%	73.23%

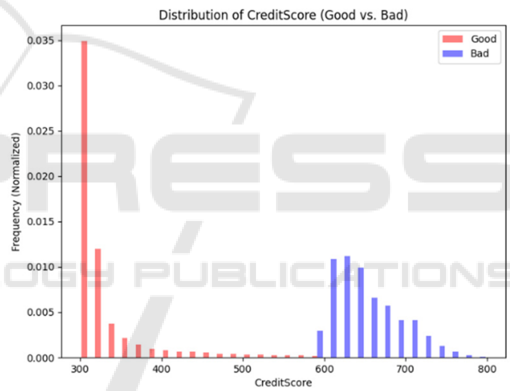

Figure 3: Distribution of CreditScore (Good xs. Bad) (Picture credit: Original).

Fig. 3 illustrates the distribution of customer credit scores, indicating that most customers have good credit status. However, there is a higher proportion of bad credit in the low segment of 300-500, and these customers tend to have higher credit risks. On the other hand, when the credit score exceeds 600, customers with good credit dominate, indicating lower credit risks. In the middle area between 500 and 600, the distribution of good and bad credit is relatively balanced, meaning moderate credit risk. This information is valuable for financial institutions to develop appropriate loan policies and interest rates for customers with different credit scores.

In summary, this analysis has revealed weak correlations among dataset variables, contributing to enhanced model accuracy. Effective data preprocessing and feature selection have been employed, and the adoption of model stacking has notably improved prediction accuracy. The distribution of customer credit scores indicates that most have good credit, with higher credit risk in the lower score range. These insights are invaluable for financial institutions when tailoring their policies and interest rates to accommodate customers with different credit profiles, thereby enhancing risk management.

5 CONCLUSION

This research endeavor embarked on a meticulous journey to construct a robust model for the evaluation of credit risk among bank customers. This involved a comprehensive fusion of data preprocessing techniques, intricate feature selection, diverse model training strategies, and the application of advanced stacking methodologies. It is noteworthy that the resultant model demonstrated not only a commendable AUC but also achieved impressive accuracy levels. Furthermore, the model possesses the unique capability to transform predicted probabilities into concrete credit scores, endowing financial institutions with vital decision-making insights of paramount significance. This symbiotic fusion of technical prowess and financial acumen forms the bedrock of this research's contributions. However, it's imperative to underscore that the enlightening power of data visualization played a pivotal role in this research, as evidenced in the intricacies of Fig. 1 and Fig. 2. These figures provided an in-depth perspective into the intricate web of inter-feature relationships and their relative significance. Likewise, the revelations encapsulated within Fig. 3 elegantly portrayed the subtleties of customer distribution across a spectrum of credit score brackets. These insights are not just enlightening; they are transformative for financial institutions. They furnish these entities with the ability to craft judicious loan policies and finely-tuned interest rate structures, thus optimizing risk management strategies. In essence, this study bequeaths a potent tool to banks and fiscal institutions, endowing them with the capacity to assess credit risks with unparalleled precision. Nonetheless, as the inexorable march of technology continues and data repositories burgeon, the immense potential persists for further refining this model. Future endeavors could delve into innovative feature engineering paradigms and leverage avant-garde modeling techniques. These forward-looking efforts would ensure that the prediction framework remains perched at the zenith of accuracy, seamlessly catering to the evolving demands of the dynamic financial sector. The horizon for advancement is boundless, and this research marks but a foundational step toward an ever-brighter future in credit risk assessment.

REFERENCES

J. X. Jiang and F. Packer, "Credit ratings of Chinese firms by domestic and global agencies: Assessing the determinants and impact," Journal of Banking & Finance, vol. 105, pp. 178-193, 2019.

M. Musdholifah, U. Hartono, and Y. Wulandari, "Banking crisis prediction: emerging crisis determinants in Indonesian banks," International Journal of Economics and Financial Issues, vol. 10, no. 2, pp. 124, 2020.

A. S. Kadam, S. R. Nikam, A. A. Aher, et al., "Prediction for loan approval using machine learning algorithm," International Research Journal of Engineering and Technology (IRJET), vol. 8, no. 04, 2021.

J. Beutel, S. List, and G. von Schweinitz, "Does machine learning help us predict banking crises?" Journal of Financial Stability, vol. 45, pp. 100693, 2019.

N. Uddin, M. K. U. Ahamed, M. A. Uddin, et al., "An Ensemble Machine Learning Based Bank Loan Approval Predictions System with a Smart Application," Available at SSRN 4376481. (This one seems like a working paper available on SSRN and not necessarily a journal paper. Thus, it might not fit perfectly into the provided format)

H. Erdal and İ. Karahanoğlu, "Bagging ensemble models for bank profitability: An empirical research on Turkish development and investment banks," Applied soft computing, vol. 49, pp. 861-867, 2016.

P. Carmona, F. Climent, and A. Momparler, "Predicting failure in the US banking sector: An extreme gradient boosting approach," International Review of Economics & Finance, vol. 61, pp. 304-323, 2019.

G. Wang, S. W. H. Kwok, D. Axford, et al., "An AUC-maximizing classifier for skewed and partially labeled data with an application in clinical prediction modeling," Knowledge-Based Systems, vol. 278, pp. 110831, 2023.

South German Credit (UPDATE). (2020). UCI Machine Learning Repository. https://doi.org/10.24432/C5QG88.