Journal of Information System Exploration and Research Vol. 2, No. 1, January 2024, pp. 11-20

Journal of Information System Exploration and Research

https://shmpublisher.com/index.php/joiser

p-ISSN 2964-1160 | e-ISSN 2963-6361

The Optimization of Credit Scoring Model Using Stacking Ensemble Learning and Oversampling Techniques

Rofik[1*], Reza Aulia[2], Khalimah Musaadah[3], Salma Shafira Fatya Ardyani[4], Ade Anggian Hakim[5]

[1]*Department of Computer Science, Faculty of Mathematics and Natural Science, Universitas Negeri Semarang, Indonesia*
[2]*Department of Science, Graduate School of Integrated Science and Technology, Shizouka Univesity, Japan*
[3]*Department of Informatics, Faculty of Engineering, Universitas Jenderal Soedirman, Indonesia*
[4]*Department of Informatics Engineering, Faculty of Computer Science, Universitas Dian Nuswantoro, Indonesia*
[5]*Department of Electrical Engineering, Faculty of Electrical and Electronic Engineering, University Tun Hussein Onn Malaysia, Malaysia*

DOI: https://doi.org/10.52465/joiser.v2i1.203
Received 27 August 2023; Accepted 30 October 2023; Available online xx December 2023

Article Info

Keywords:
Credit scoring;
Stacking ensemble learning;
XGBoost;
SMOTE

Abstract

Credit risk assessment plays an important role in efficient and safe banking decision-making. Many studies have been conducted to analyze credit scoring with a focus on achieving high accuracy. However, predicting credit scoring decisions also requires model construction that handles class imbalance and proper model implementation. This research aims to increase the accuracy of credit assessment by balancing data using Synthetic Minority Oversampling (SMOTE) and applying ensemble stacking learning techniques. The proposed model utilizes a base learner consisting of Random Forest, SVM, Extra-Tree Classifier, and XGboost as a meta-learner. Then to handle unbalanced classes using SMOTE. The research process was carried out in several stages, namely Data Collection, Preprocessing, Oversampling, Modeling, and Evaluation. The model was tested using the German Credit dataset by applying cross-validation. The evaluation results show that the stacking ensemble learning model developed has optimal performance, with an accuracy of 83.21%, precision of 79.29%, recall of 91.78%, and f1-score of 85.08%. This research shows that optimizing the stacking ensemble learning model with data balancing using SMOTE in credit scoring can improve performance in credit scoring.

1. Introduction

In the current era of information technology development and financial innovation, the credit scoring process has become a crucial cornerstone in the decision-making of banks and lending

[*] ***Corresponding Author:***

Rofik,
Department of Computer Science,
Faculty of Mathematics and Natural Science, Universitas Negeri Semarang,
Semarang, Indonesia.
Email: rofikn4291@students.unnes.ac.id

institutions [1]. Credit scoring is a job that aims to assess the credit risk of a prospective customer or business entity, which helps financial institutions make efficient and accurate decisions regarding the granting of credit [2]. In applying data analysis techniques and predictive models, credit scoring plays an important role in maintaining the balance between safe financing and lending to the right parties. This is because inappropriate lending decisions can result in huge losses [2].

Lending behavior and technological developments have changed the landscape of the credit assessment process [3]. Especially in this digital era, online lending has become a popular alternative for individuals and businesses to obtain funds without involving traditional financial institutions such as banks [4]. Such loans utilize the convenience and transparency of online platforms, resulting in a faster and easier process compared to conventional approaches [5]. However, online consumer lending is of great concern to the Company, as online consumer lending typically carries a higher credit risk than businesses in the conventional lending system [6].

Based on information from the Wind economic database in 2019, in China, the accumulated rate of 1-year payment defaults of some lending institutions or securities backing online consumer loans even reached 9% [7]. Although it is difficult to ensure that a borrower will default in the future, with advanced analytical approaches and appropriate modeling techniques, credit scoring can indicate potential default risk before a credit transaction is made [8], [9].

However, this development also brings about new challenges in credit risk assessment. In addition to the disadvantages of the traditional approach that can be easily affected by sample selection bias, as it only uses a sample of accepted applicants, while the applicant population also includes rejected applicants [10], [11]. The existence of large and highly variable data, as well as the complexity of factors that can affect credit decisions, makes demands on lending institutions to still be able to perform credit scoring quickly and accurately [12]. Therefore, many researchers have developed techniques to assess credit risk with data mining [11], [13]–[16]. Data mining is a machine learning method that aims to extract valuable information from existing data [3]. Using this technique, important information can be identified from customer data, credit history, and other relevant factors.

There have been many studies that implement data mining in credit scoring. Research by [17] who did a new approach to assessing credit applications by giving a binary score, by combining a Genetic Algorithm (GA) with Support Vector Machine (SVM). By applying 2 levels, which are determining the SVM parametrization and finding the most weighted feature set, this research was able to achieve an accuracy of 80.70%. Olivares et al in their research [18] explored the application of discrete-time joint models in credit scoring. The study combined survival analysis with longitudinal data by integrating variable covariates in survival analysis. From the study, it was found that the inclusion of time-varying covariates in the survival model improved the prediction of credit scoring. Using Australian, German, and Japanese datasets, the research in [19] focuses on the implementation of the Extreme Learning Machine (ELM) classification tool for the credit scoring analysis model. Since ELM requires more hidden neurons and random determination of input weights and hidden biases, the study proposed a novel activation function and evolutionary approach to obtain optimized weights and biases using the Bat algorithm. From the model that has been built, this research can achieve consecutive accuracy of the Australian, German, and Japanese datasets of 89.92%, 81.18%, and 88.35%. The research conducted by [12] is focused on a new development called soft reordering one-dimensional CNN (SR-1D-CNN) which is designed to adaptively restructure the original table data to better suit CNN learning. By using 5 datasets from Polish, Ashare, GiveMeSomeCredit, Lending Club and HomerCreditDefaultRisk, the model built was able to produce the greatest accuracy of 95.18 from the Polish dataset.

The research conducted by the author will focus on developing a credit-scoring model using stacking ensemble learning techniques and handling unbalanced data. Despite technological advances and the application of machine learning models in credit scoring analysis, the main problem that often arises is the inability of the model to explain predictions and data imbalance [20]. As done in [9], [21]–[23] which focuses on handling class imbalance. It can be concluded that the performance of the model decreases inversely with increasing the level of class imbalance. This is proven by [1] by evaluating class imbalance using LIME and SHAP stability. This shows that the resulting interpretations of LIME and SHAP are less stable as class imbalance increases, concluding that class imbalance does hurt machine learning interpretations. Research conducted by [24]–[26] also revealed that ensemble models tend to produce good performance. Research [27] also reveals that in the context of credit scoring, ensemble methods based on decision trees such as random forest algorithms produce better classification performance compared to standard logistic regression models.

2. Method

In this research, credit scoring analysis is carried out with several stages, namely, data collection, preprocessing, oversampling, modeling, and evaluation. The framework of the stages of this research can be seen in Figure 1.

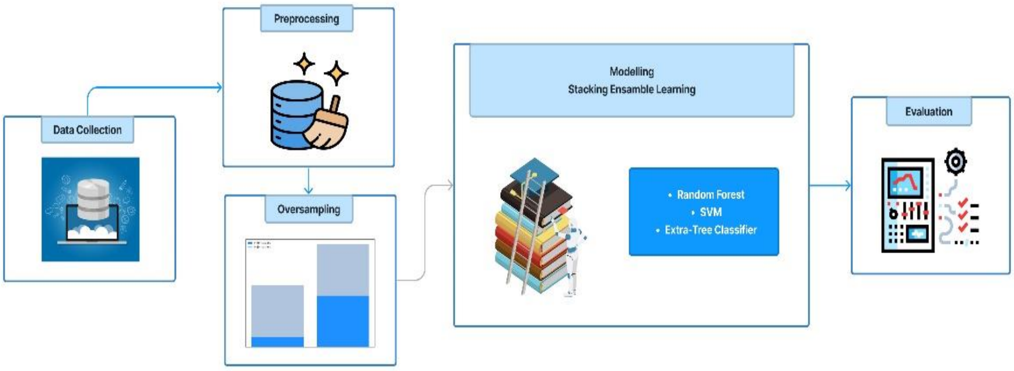


Figure 1. Research Framework of Credit Scoring

A comprehensive framework is proposed as a guide to understanding the process that must be done to achieve high-accuracy credit scoring prediction results. The framework incorporates various important stages starting from data collection to evaluation. Figure 1 shows the main steps taken in this research. Through these steps, it is expected to present a solution that can overcome the problems that usually arise in credit scoring including data imbalance. A more detailed explanation of each stage in this research framework can be seen as follows.

2.1 Data Collection

In this section, data related to customer credit is grouped. The dataset used in this research is a German dataset sourced from UCI Machine Learning. The dataset consists of 20 attributes, i.e. "status", "duration", "credit_history", "purpose", "amount", "savings", "tenure", "installment_amount", "installment_level", "status_gender", "debtor_other", "current_place_of_residence", "property", "age", "installment plan_other", "place_of_residence", "amount_of_credit", "occupation", "amount_of_responsibility", "telephone", "foreign_worker", and "credit_risk" as the labels. The dataset consists of 1000 records of credit data. Of these data, 700 records are good classes and 300 are bad classes. The dataset used in this research can be accessed via the URL link: https://archive.ics.uci.edu/dataset/573/south+german+credit+update.

2.2 Data preprocessing

Data preprocessing is used to clean and prepare the dataset so that it can be implemented and support the modeling stage. Data separation is done to separate features (X) which are features in the credit scoring dataset, and target variables (Y) which are class data. Feature encoding is also done using One-Hot Encoding. Data standardization is also done using Standard Scaler. This is done to make all data in each feature have the same scale and avoid model sensitivity issues due to scale differences.

2.3 Data oversampling

Oversampling of data is done to overcome datasets that have unbalanced classes [28]. Where the number of good consumers is more than the number of bad consumers. In this case are accepted credit applicants and rejected credit applicants, to overcome this problem which will have a habitual impact on the model, the oversampling technique is carried out [29]. With the application of oversampling, the model built will not be more inclined to the majority class only during the training process but is good at generalizing both classes. The oversampling technique performed in this research is the Synthetic

Minority Over-sampling Technique Evaluation (SMOTE) method. This technique is capable of generating synthesized data [30]. The sample is generated by increasing the sample that is different from the minority class sample [31]. SMOTE is a data augmentation method technique for classification datasets, which improves recognition performance without increasing the risk of data leakage [32].

2.4 Modeling with Stacking Ensemble Learning.

Initialization of the base model is done by selecting the base model used. In this research, Random Forest, SVM, and Extra-Tree Classifier algorithms were used. Some of the reasons for choosing these algorithms as base models include Random Forest, with its ability to overcome overfitting [33] and can produce stable and accurate predictions, which is used to combine several decision trees to produce the final prediction. Where the prediction results are drawn through voting from decision trees that work independently. It also outperforms logistic regression algorithms and is developing into a major algorithm in the credit scoring sector [34]. The SVM algorithm is also chosen and applied as a base model, due to its popularity and efficiency in solving classification and regression problems [35]. The working concept of this algorithm is to separate two classes by maximizing the margin between the two classes. Meanwhile, the Extra-Tree Classifier is also applied as a base model because in some studies it has been proven to show good performance [36]. This algorithm belongs to the ensemble learning category which is similar to Random Forest. However, this algorithm works by building a decision tree with random features and subsampling. Its advantage is that it is suitable for large data and its computational speed is also high.

These algorithms are trained using training data and can understand patterns in the data and make predictions. Once trained, the algorithms are used to predict test data, where each model predicts the probability for each class. The probability results from each base model are used by the final estimator or in this study, XGBoost. The ensemble uses the probability results from the base models as input features to help the meta-model produce better final predictions.

Previously the data was divided into training data and testing data. The ratio is 80% training data and 20% testing data. A cross-validation of 5 times was performed on the ensemble stacking model developed using the 80% training data. This method is done because it has a good impact on the model built [37]. The concept is to train the model by using alternating data as training and testing, which from this process can help the model generalize data that is not seen in training. The prediction results provide a probability value for each credit risk class which is then used to determine the final decision. From these probability values, various thresholds that separate high and low credit risk classes can be explored. This research also tested several thresholds from the range of 0.1 to 0.9 in cross-validation. Threshold testing is done in a separate loop after all cross-validation iterations are completed. The threshold that yields the greatest accuracy among other threshold tests is used to classify the prediction results.

2.5 Evaluation

The confusion metric is one of the evaluation metrics used to analyze the performance of the stacking ensemble learning model, which is also used in this study. The confusion matrix can provide an overview of the prediction and actual state given by the algorithm model. Confusion Matrix has 4 important elements. Among them is True Positives (TP), which represents the amount of data that is actually in the positive class and predicted by the model as a positive class. Then True Negative (TN), which reflects the amount of data that is actually in the negative class and is predicted by the model as a negative class. While False Positive (FP) is the amount of data that is actually in the negative class but is predicted by the model as a positive class. False Negative (FN) is the amount of data that is actually in the positive class but predicted by the model as a negative class. By using the Confusion matrix, we can calculate and evaluate the performance of the model that has been built through metrics such as accuracy, precision, recall, and f1-score. The calculation details can be seen below.

1. Accuracy

Accuracy is a value that indicates how accurate the model is in predicting the entire data. Measured by the formula:

$$Accuracy = \frac{(TP + TN)}{(TP + TN + FP + FN)} \tag{1}$$

2. Precision
Precision measures the degree to which work predicted to be fake is fake. We can calculate it with the following formula:

$$Precision = \frac{TP}{(TP + FP)} \quad (2)$$

3. Recall (Sensitivity)
Recall measures the extent to which the model successfully detects fake jobs overall. Recall calculated by the following formula:

$$Recall = \frac{TP}{(TP + FN)} \quad (3)$$

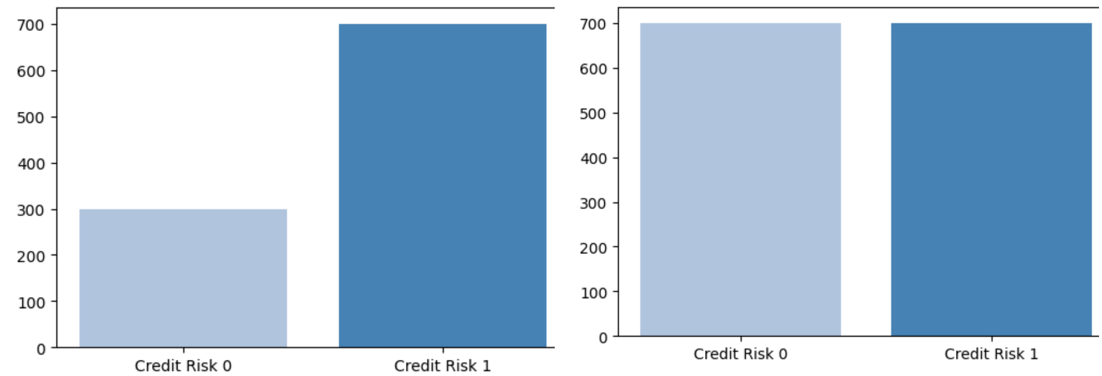
4. F1-Score
The F1-Score is a combination of precision and recall into a single metric that yields the overall model performance. This can be calculated by the following formula:

$$F1 - Score = \frac{2 * (Presisi * Recall)}{(Presisi + Recall)} \quad (4)$$

3. Results and Discussion

This research uses the German credit dataset that has been used in previous studies. The research was carried out in stages starting with data collection, namely collecting datasets where datasets used were obtained from UCI Machine Learning. Perform preprocessing, separating features (X) and target variables (Y). Perform data coding using One-Hot Encoding, and standardize data using StandardScaler. Then oversampling the data using SMOTE to overcome data imbalances that can cause the model's performance to be not optimal in classification. The modeling stage is carried out to build a stacking ensemble learning model that combines Random Forest, SVM, and Extra-Tree Classifier algorithms as base learner. And Xgboost as a meta-learner model. The next stage is evaluation, cross-validation is done by calculating the probability of prediction using the stacking model. Apply to determine the Best Threshold to maximize accuracy by iterating through several threshold values. Whichever threshold is best is used to classify the data and calculate cross-validation accuracy. Then the model performance is calculated through accuracy, precision, recall, and f1-score. The following are the threshold experiments performed on cross-validation that correlate with the accuracy obtained.

The results of oversampling to balance the data using the SMOTE method against the amount of data can be seen in Figure 2.



(a) (b)

Fig 2. Data distribution (a) before oversampling and (b) after oversampling.

Handling unbalanced classes is done using the SMOTE method which augments the data from the minority class, resulting in data that is balanced between the classes. This reduces bias in predictions and allows the model to learn better patterns from both classes in the dataset. As a result of the SMOTE implementation, there is a class balance of 700, whereas previously the bad credit risk class was 300 and the good credit risk class was 700.

A heatmap is displayed that shows the correlation between the features in the dataset and its target, which in this case is 'credit_risk'. The darker the color, the stronger the correlation, and vice versa. The value that shows the size of the correlation if it is close to 1 means a positive correlation, while otherwise it is negatively correlated. Whereas when the value is 0, it means that there is no correlation between the two features.

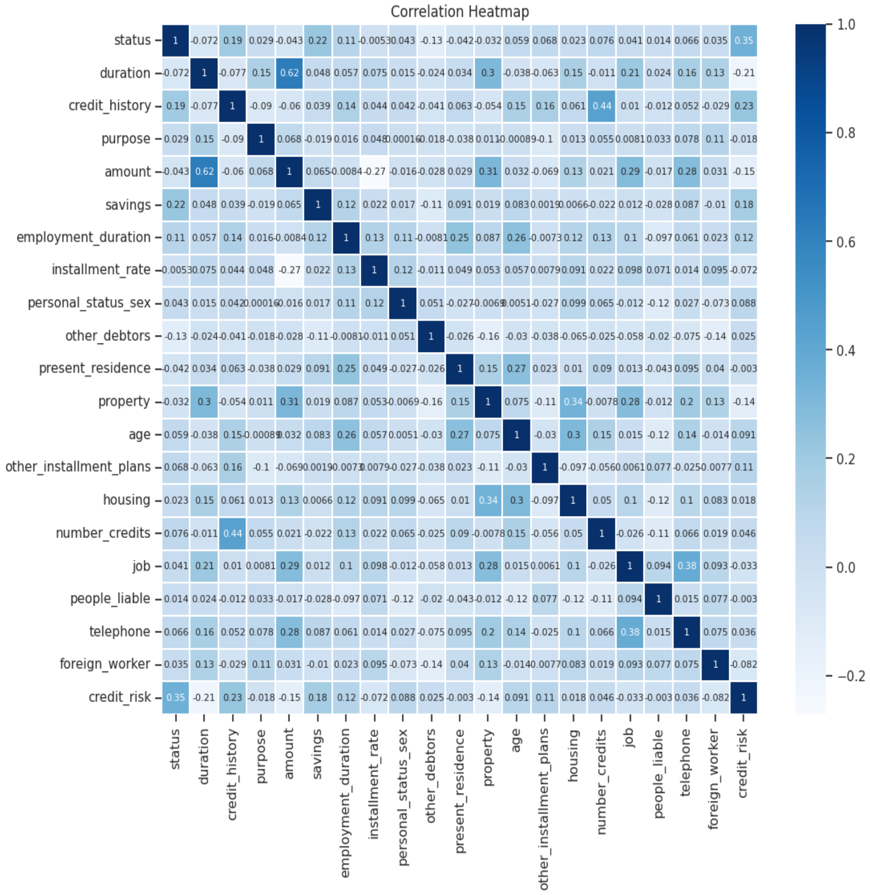


Figure 3. Heatmap of correlation between features

The 3 features that are strongly positively correlated with the target include status (checking account status), credit history, and savings. The most negatively correlated features include duration (length of loan), amount (amount of money borrowed), and property (cars, real property, buildings, and so on). The experimental analysis of several thresholds from 0.1 to 0.9 on the credit score prediction model using the stacking ensemble learning technique can be seen in Figure 4.

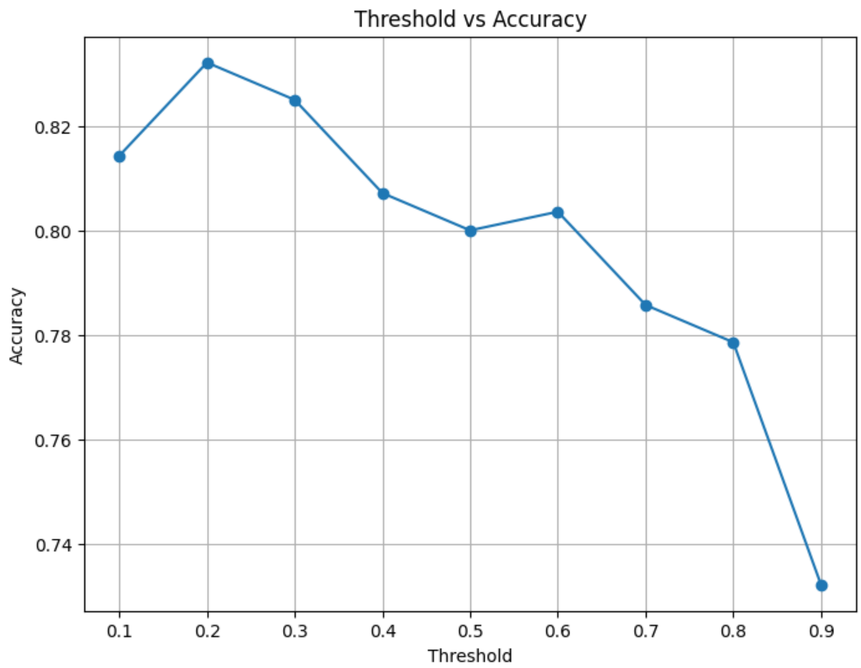


Figure 4. Model performance on testing each threshold

In each cross-validation fold, the model is fit to the training which is then used to predict the probability of a positive class. Different thresholds are also applied. From the figure, it can be seen that the average accuracy changes as the threshold moves from 0.1 to 0.9. Each point on the grid represents the average accuracy of the model across all cross-validation folds at a particular threshold. The largest average accuracy obtained was 83.21% with a threshold of 0.2.

This research uses a stacking ensemble learning model by combining 3 algorithms, namely Random Forest, SVM, and Extra-Tree Classifier. The results obtained state that the model built successfully produces good performance and can improve the accuracy performance of previous research. The test results were evaluated using a confusion matrix in the form of accuracy, precision, recall, and score. The model built was able to produce the greatest accuracy of 83.21%, precision of 79.29%, recall of 91.78%, and F1-score of 85.07%. A comparison of the performance of the ensemble stacking model results built in this study with previous research models can be seen in Table 1.

Table 1. Comparison of techniques and results from previous research

Approach	Techniques or algorithms used	Year	Accuracy
[38]	Combines the benefits of feature selection and ensemble frameworks. Using 5 basic classification algorithms and combining them with the weighted voting approach	2018	77.12%
[39]	Bayesian optimization and PSO implementation	2019	78.30%
[17]	Combination of genetic algorithm and SVM	2020	80.70%
[40]	Multi-grained augmented gradient boosting decision trees (mg-GBDT), (Gradient boosting decision trees, Multi-grained scanning)	2021	77.15%
Proposed Method	Optimization of Credit Scoring Analysis Model using Stacking Ensemble Learning Approach and Oversampling with SMOTE	2023	83.21%

4. Conclusion

In this research, credit scoring classification between good credit risk and bad credit risk is carried out using stacking ensemble learning from Random Forest, SVM, and Extra-Tree Classifier algorithms.

The meta-learner model is XGBoost. This research shows the effectiveness of the stacking ensemble learning model in classifying good and bad credit risk, namely between accepted and unaccepted credit applicants. Oversampling using SMOTE is used to overcome unbalanced datasets. An inter-class boundary search was also conducted using thresholds from 0.1 to 0.9 for classification. The evaluation results showed that the ensemble stacking model successfully improved the performance in distinguishing good and bad credit risk. The resulting performance of the stacking model achieved the best accuracy of 83.21% with a precision of 79.28%, recall of 91.78%, and f1-score of 85.97%. And the best threshold to separate the two classes is 0.2. In future research, it is recommended to further explore larger data, perform feature selection, and try to create new models to achieve more optimal credit scoring performance.

References

[1] Y. Chen, R. Calabrese, and B. Martin-Barragan, “Interpretable machine learning for imbalanced credit scoring datasets,” *Eur. J. Oper. Res.*, no. xxxx, 2023, doi: 10.1016/j.ejor.2023.06.036.

[2] Z. Zhang, Y. Li, Y. Liu, and S. Liu, “A local binary social spider algorithm for feature selection in credit scoring model,” *Appl. Soft Comput.*, vol. 144, p. 110549, 2023, doi: 10.1016/j.asoc.2023.110549.

[3] M. A. Muslim *et al.*, “New model combination meta-learner to improve accuracy prediction P2P lending with stacking ensemble learning,” *Intell. Syst. with Appl.*, vol. 18, no. December 2022, p. 200204, 2023, doi: 10.1016/j.iswa.2023.200204.

[4] X. Ma, J. Sha, D. Wang, Y. Yu, Q. Yang, and X. Niu, “Study on a prediction of P2P network loan default based on the machine learning LightGBM and XGboost algorithms according to different high dimensional data cleaning,” *Electron. Commer. Res. Appl.*, vol. 31, no. February, pp. 24–39, 2018, doi: 10.1016/j.elerap.2018.08.002.

[5] X. Fu, S. Zhang, J. Chen, T. Ouyang, and J. Wu, “A Sentiment-Aware Trading Volume Prediction Model for P2P Market Using LSTM,” *IEEE Access*, vol. 7, pp. 81934–81944, 2019, doi: 10.1109/ACCESS.2019.2923637.

[6] M. Di Maggio and V. Yao, “Fintech Borrowers: Lax Screening or Cream-Skimming?,” *Rev. Financ. Stud.*, vol. 34, no. 10, pp. 4565–4618, 2021, doi: 10.1093/rfs/hhaa142.

[7] Y. Xia, Y. Li, L. He, Y. Xu, and Y. Meng, “Incorporating multilevel macroeconomic variables into credit scoring for online consumer lending,” *Electron. Commer. Res. Appl.*, vol. 49, no. March, p. 101095, 2021, doi: 10.1016/j.elerap.2021.101095.

[8] C. Bai, B. Shi, F. Liu, and J. Sarkis, “Banking credit worthiness: Evaluating the complex relationships,” *Omega (United Kingdom)*, vol. 83, pp. 26–38, 2019, doi: 10.1016/j.omega.2018.02.001.

[9] F. Shen, X. Zhao, G. Kou, and F. E. Alsaadi, “A new deep learning ensemble credit risk evaluation model with an improved synthetic minority oversampling technique,” *Appl. Soft Comput.*, vol. 98, p. 106852, 2021, doi: 10.1016/j.asoc.2020.106852.

[10] Y. Kang, N. Jia, R. Cui, and J. Deng, “A graph-based semi-supervised reject inference framework considering imbalanced data distribution for consumer credit scoring,” *Appl. Soft Comput.*, vol. 105, p. 107259, 2021, doi: 10.1016/j.asoc.2021.107259.

[11] R. A. Mancisidor, M. Kampffmeyer, K. Aas, and R. Jenssen, “Deep generative models for reject inference in credit scoring,” *Knowledge-Based Syst.*, vol. 196, p. 105758, 2020, doi: 10.1016/j.knosys.2020.105758.

[12] H. Qian, P. Ma, S. Gao, and Y. Song, “Soft reordering one-dimensional convolutional neural network for credit scoring,” *Knowledge-Based Syst.*, vol. 266, p. 110414, 2023, doi: 10.1016/j.knosys.2023.110414.

[13] Y. Wang, Y. Jia, Y. Zhong, J. Huang, and J. Xiao, “Balanced incremental deep reinforcement learning based on variational autoencoder data augmentation for customer credit scoring,” *Eng. Appl. Artif. Intell.*, vol. 122, no. March, p. 106056, 2023, doi: 10.1016/j.engappai.2023.106056.

[14] H. He, Z. Wang, H. Jain, C. Jiang, and S. Yang, “A privacy-preserving decentralized credit scoring method based on multi-party information,” *Decis. Support Syst.*, vol. 166, no. November 2022, p. 113910, 2023, doi: 10.1016/j.dss.2022.113910.

[15] Y. Wu, W. Huang, Y. Tian, Q. Zhu, and L. Yu, “An uncertainty-oriented cost-sensitive credit

scoring framework with multi-objective feature selection," *Electron. Commer. Res. Appl.*, vol. 53, no. March, p. 101155, 2022, doi: 10.1016/j.elerap.2022.101155.
[16] D. M. B. Silva, G. H. A. Pereira, and T. M. Magalhães, "A class of categorization methods for credit scoring models," *Eur. J. Oper. Res.*, vol. 296, no. 1, pp. 323–331, 2022, doi: 10.1016/j.ejor.2021.04.029.
[17] D. Şen, C. Ç. Dönmez, and U. M. Yıldırım, "A Hybrid Bi-level Metaheuristic for Credit Scoring," *Inf. Syst. Front.*, vol. 22, no. 5, pp. 1009–1019, 2020, doi: 10.1007/s10796-020-10037-0.
[18] V. Medina-Olivares, R. Calabrese, J. Crook, and F. Lindgren, "Joint models for longitudinal and discrete survival data in credit scoring," *Eur. J. Oper. Res.*, vol. 307, no. 3, pp. 1457–1473, 2023, doi: 10.1016/j.ejor.2022.10.022.
[19] D. Tripathi, D. R. Edla, V. Kuppili, and A. Bablani, "Evolutionary Extreme Learning Machine with novel activation function for credit scoring," *Eng. Appl. Artif. Intell.*, vol. 96, no. February, p. 103980, 2020, doi: 10.1016/j.engappai.2020.103980.
[20] X. Dastile, T. Celik, and M. Potsane, "Statistical and machine learning models in credit scoring: A systematic literature survey," *Appl. Soft Comput. J.*, vol. 91, p. 106263, 2020, doi: 10.1016/j.asoc.2020.106263.
[21] Y. Li, T. Bellotti, and N. Adams, "Issues Using Logistic Regression With Class Imbalance, With a Case Study From Credit Risk Modelling," *Found. Data Sci.*, vol. 1, no. 4, pp. 389–417, 2019, doi: 10.3934/fods.2019016.
[22] K. Zhang *et al.*, "Label correlation guided borderline oversampling for imbalanced multi-label data learning," *Knowledge-Based Syst.*, vol. 279, p. 110938, 2023, doi: 10.1016/j.knosys.2023.110938.
[23] X. Tao, X. Guo, Y. Zheng, X. Zhang, and Z. Chen, "Self-adaptive oversampling method based on the complexity of minority data in imbalanced datasets classification," *Knowledge-Based Syst.*, vol. 277, p. 110795, 2023, doi: 10.1016/j.knosys.2023.110795.
[24] G. E-mail, K. W. De Bock, K. Coussement, S. Lessmann, and K. W. De Bock, "Version of Record: https://www.sciencedirect.com/science/article/pii/S0377221720300898," pp. 1–39.
[25] W. Yin, B. Kirkulak-Uludag, D. Zhu, and Z. Zhou, "Stacking ensemble method for personal credit risk assessment in Peer-to-Peer lending," *Appl. Soft Comput.*, vol. 142, p. 110302, 2023, doi: 10.1016/j.asoc.2023.110302.
[26] P. Sulikowski and T. Zdziebko, "Churn factors identification from real-world data in the telecommunications industry: Case study," *Procedia Comput. Sci.*, vol. 192, pp. 4800–4809, 2021, doi: 10.1016/j.procs.2021.09.258.
[27] E. Dumitrescu, S. Hué, C. Hurlin, and S. Tokpavi, "Machine learning for credit scoring: Improving logistic regression with non-linear decision-tree effects," *Eur. J. Oper. Res.*, vol. 297, no. 3, pp. 1178–1192, 2022, doi: 10.1016/j.ejor.2021.06.053.
[28] W. F. Abror and M. Aziz, "Journal of Information System Bankruptcy Prediction Using Genetic Algorithm-Support Vector Machine (GA-SVM) Feature Selection and Stacking," vol. 1, no. 2, pp. 103–108, 2023.
[29] J. Mushava and M. Murray, "A novel XGBoost extension for credit scoring class-imbalanced data combining a generalized extreme value link and a modified focal loss function," *Expert Syst. Appl.*, vol. 202, no. March, p. 117233, 2022, doi: 10.1016/j.eswa.2022.117233.
[30] A. Fernández, S. García, F. Herrera, and N. V. Chawla, "SMOTE for Learning from Imbalanced Data: Progress and Challenges, Marking the 15-year Anniversary," *J. Artif. Intell. Res.*, vol. 61, pp. 863–905, 2018, doi: 10.1613/jair.1.11192.
[31] J. Sun, J. Lang, H. Fujita, and H. Li, "Imbalanced enterprise credit evaluation with DTE-SBD: Decision tree ensemble based on SMOTE and bagging with differentiated sampling rates," *Inf. Sci. (Ny).*, vol. 425, pp. 76–91, 2018, doi: 10.1016/j.ins.2017.10.017.
[32] A. Imakura, M. Kihira, Y. Okada, and T. Sakurai, "Another use of SMOTE for interpretable data collaboration analysis," *Expert Syst. Appl.*, vol. 228, no. January, 2023, doi: 10.1016/j.eswa.2023.120385.
[33] A. Hoarau, A. Martin, J. C. Dubois, and Y. Le Gall, "Evidential Random Forests," *Expert Syst. Appl.*, vol. 230, no. February, p. 120652, 2023, doi: 10.1016/j.eswa.2023.120652.
[34] Y. Liu, M. Yang, Y. Wang, Y. Li, and T. Xiong, "Applying machine learning algorithms to predict default probability in the online credit market: Evidence from China," *Int. Rev. Financ. Anal.*, vol. 79, no. November 2021, p. 101971, 2022, doi: 10.1016/j.irfa.2021.101971.
[35] M. A. Ganaie, A. Kumari, A. Girard, J. Kasa-Vubu, and M. Tanveer, "Diagnosis of Alzheimer's

disease via Intuitionistic fuzzy least squares twin SVM," *Appl. Soft Comput.*, vol. 149, no. September, 2023, doi: 10.1016/j.asoc.2023.110899.

[36] H. F. Chen *et al.*, "Predicting residual stress of aluminum nitride thin-film by incorporating manifold learning and tree-based ensemble classifier," *Mater. Chem. Phys.*, vol. 295, no. 300, p. 127070, 2023, doi: 10.1016/j.matchemphys.2022.127070.

[37] S. M. Malakouti, "Improving the prediction of wind speed and power production of SCADA system with ensemble method and 10-fold cross-validation," *Case Stud. Chem. Environ. Eng.*, vol. 8, no. April, p. 100351, 2023, doi: 10.1016/j.cscee.2023.100351.

[38] D. Tripathi, D. R. Edla, V. Kuppili, A. Bablani, and R. Dharavath, "Credit Scoring Model based on Weighted Voting and Cluster based Feature Selection," *Procedia Comput. Sci.*, vol. 132, no. Iccids, pp. 22–31, 2018, doi: 10.1016/j.procs.2018.05.055.

[39] S. Guo, H. He, and X. Huang, "A Multi-Stage Self-Adaptive Classifier Ensemble Model With Application in Credit Scoring," *IEEE Access*, vol. 7, pp. 78549–78559, 2019, doi: 10.1109/ACCESS.2019.2922676.

[40] W. Liu, H. Fan, and M. Xia, "Step-wise multi-grained augmented gradient boosting decision trees for credit scoring," *Eng. Appl. Artif. Intell.*, vol. 97, no. May 2020, p. 104036, 2021, doi: 10.1016/j.engappai.2020.104036.