Scientific Journal of Informatics
Vol. 11, No. 4, Nov 2024

p-ISSN 2407-7658 https://journal.unnes.ac.id/journals/sji/index e-ISSN 2460-0040

A Performance Comparison of Data Balancing Model to Improve Credit Risk Prediction in P2P Lending

Dwika Ananda Agustina Pertiwi[1*], Kamilah Ahmad[2], Jumanto Unjung[3], Much Aziz Muslim[4]
[1,2]Department of Technology Management, Universiti Tun Hussein Onn Malaysia, Malaysia
[3,4]Department Computer Science, Universitas Negeri Semarang, Indonesia

Abstract.
Purpose: The problem of imbalanced datasets often affects the performance of classification models for prediction, one of which is credit risk prediction in P2P lending. To overcome this problem, several data balancing models have been applied in the existing literature. However, existing research only evaluates performance based on classification model performance. Thus, in addition to measuring the performance of classification models, this study involves the contribution of the performance of data balancing models including Random Oversampling (ROS), Random Undersampling (RUS), and Synthetic Minority Oversampling (SMOTE).
Methods: This research uses the Lending Club dataset with an imbalanced ratio (IR) of 4.098, and 2 classifiers such as LightGBM and XGBoost, as well as 10 cross-validation to assess the performance of the data balancing model including Random Oversampling (ROS), Random Undersampling (RUS), and Synthetic Minority Oversampling (SMOTE). Then the model is evaluated using the metrics of accuracy, recall, precision, and F1-score.
Result: The research results show that SMOTE has superior performance as a data balancing model in P2P lending, with an accuracy of the LightGBM+SMOTE model of 92.56% and the XGBoost+SMOTE model of 92.32%, where this performance is better than other models.
Novelty: This research concludes that SMOTE as a data balancing model to improve credit risk prediction in P2P lending has superior performance. Apart from that, in this case, we find that the larger the data size used as a model training sample, the superior performance obtained by the classification model in predicting credit risk in P2P lending.

Keywords: P2P lending, Data balancing model, LightGBM, XGBoost
Received September 2024 / **Revised** November 2024 / **Accepted** November 2024

INTRODUCTION

Peer-to-peer (P2P) lending is a financial technology business model [1]. P2P lending is a method of obtaining credit without the involvement of financial institutions like banks in the decision-making process. It has a greater chance of obtaining favorable terms than the conventional banking system [2]. P2P lending was created in 2005, and it has lately experienced significant global growth. The rise in P2P loans is a result of banks' improved credit standing [3]. As a result, a short-term credit history with a poor credit rating may indicate possible credit risk. The lenders need to be better able to assess credit risk as P2P lending becomes increasingly common as risks arise [3]. This is most likely related to the use of data mining techniques, specifically machine learning algorithms and alternative data sources.

In addition, to being a helpful tool for lending institutions when deciding whether to approve credit applications, a strong model may also educate clients about actions that could lower their credit ratings [4]. Utilizing financial data, such as exchange records, client transactions, company transactional data, and so forth, to forecast client business success or individual credit card data and reduce loss and susceptibility is the main driving force behind risk prediction. Several risk prediction models, such as Light Gradient Boosting Machine (LightGBM) [5] and Extreme Gradient Boosting (XGBoost) [6], are based on tree classifier techniques.

Lending Club is one of the largest online marketplaces for peer-to-peer lending in the United States. As noted on its website, there is potential for investors to diversify their portfolios because the market for consumer loans has topped $3 trillion. Most of the information concerning loan requests is also shared, in addition to the findings that P2P lending organizations make regarding the creditworthiness of potential

*Corresponding author.
Email addresses: hp220072@student.uthm.edu.my (Pertiwi)
DOI: 10.15294/sji.v11i4.14018

borrowers. These data may be used for a variety of purposes, including seeing how groups of users have behaved over time and between loans, but their greatest value lies in their ability to generalize user behavior and teach computational techniques for automatic credit assessment. Machine learning techniques are frequently used in this context to develop creditworthiness models for lending clients, and they have demonstrated respectable success rates [7], [8], [9]. However, one of the main drawbacks of these models is that there is an imbalanced ratio between the classes, meaning that there are many more creditworthy clients than non-creditworthy ones [10].

According to Kaur & Gosain in 2018, the P2P lending data set has the characteristics of imbalanced data classes, and the facts show that the classes of fully repaid loans and defaulted loans are not the same. This problem will make it difficult to make predictions using highly imbalanced data sets because the classifier tends to identify the majority class over the minority class, and these imbalanced data classes can reduce the performance of the classification model. Therefore, addressing the problem of categorization of imbalanced data sets in these situations is crucial [11]. Various strategies have been suggested in the machine learning literature to address class imbalance, but few have been used in the context of P2P lending. This study compares undersampling and oversampling learning to handle imbalanced data sets. Meanwhile, for the machine learning scheme, we use LightGBM and XGBoost to classify the default risk of P2P lending.

This research paper is organized into the following sections. The proposed research framework is described in the 'Research Methodology' section. Experimental results and discussion are presented in the 'Results and Discussion' section part. Conclusions and recommendation directions are discussed in the 'Conclusions' section.

RELATED WORKS

The study of default risk in P2P lending platforms, is an area that has drawn the attention of several academics, including Yang et al., in 2019 [12]; Swee et al., in 2022 [13]; Rao et al., in 2020 [14]; Teply & Polena in 2020 [15]; and Zhu in 2019 [16]. Peer-to-peer lending has gained popularity as a replacement for conventional financial institutions. Since most middle-class people lost their ability to qualify for loans after the 2008 financial crisis, P2P lending has become the preferred way for many people to get loans [17]. According to Ge et al., in 2017, information asymmetry has developed since lenders and borrowers in an online P2P lending process do not physically contact. Therefore, creating an effective and trustworthy credit risk assessment technique to lower investment risk without requiring human interaction is crucial for the continuing expansion of the P2P lending industry [18].

According to Bao et al., in 2019, credit scoring is a crucial instrument for classifying borrowers' creditworthiness for financial institutions [19]. He thinks that successful applications of machine learning algorithms to credit scoring have been made; this study tests the proposed method using P2P credit datasets. The proposed method combines machine learning algorithms, including classification and clustering algorithms, and the results demonstrate better performance than a single classifier alone. In the meanwhile, [20] and Bachmann et al. in 2011 examined the history of P2P lending and contrasted its benefits and drawbacks. Then they described the principles behind P2P lending and contrasted it with conventional bank lending. Then, Serrano-Cinca et al., in 2015 analyzed many characteristics to predict the default risk of P2P lending using statistical techniques such as Pearson's correlation, point biserial correlation, and chi-square test. They used various 7 features to generate 7 logistic regression models for determining the most significant predictor of default [22].

The aim of Cai & Zhang in 2020 is to develop an evaluation model for choosing borrowers by reducing higher risk [23]. A credit risk prediction model for peer-to-peer lending was created using the data mining approach. This study uses the Lending Club dataset from 2018 and an exploratory data analysis approach to focus on categorization while taking the significance of variables into account. In 2019, Zhu et al., in used data from Lending Club to compare the classification system for assessing the likelihood of default on P2P loans. They compare the SVM, decision tree, logistic regression, and random forest algorithms in their research [24].

Finally, Vinod et al., in 2016 analyzed the Lending Club data set to confirm which variables were critical and identify which people had a higher likelihood of timely and interest-bearing debt repayment [25]. They concluded that random forest is the most suitable classifier to detect which debtors would not pay their bills on time, while a single decision tree was the best for identifying creditworthy consumers, using precision and accuracy as performance indicators.

Despite the effort put forth in the aforementioned studies the data set class imbalance feature has all but been completely ignored. When there are significantly more instances of one class (the majority class) than another (the minority class), the data is said to be imbalanced. An unusual instance of an- order problem is when a dataset is imbalanced and the distribution of classes is not equal. In imbalanced data sets, such as the majority class and the minority class. Such data present a problem for data mining since traditional classification methods often take into account a balanced training set, which implies a bias towards the majority class [26]. On imbalanced data sets, in research by Chawla et al., in 2002 observed the synthetic minority oversampling technique (SMOTE), the ensemble-based method, and the SMOTE- Boost method. In SMOTE topic, the majority class sample contributes more to creating the synthetic neighbors, which may improve the performance of SMOTE [27]. Additionally, in Chawal et al., in 2016 reviewed the problem of imbalanced datasets and various research improvements [28].

METHODS

This segment shows the experimental steps for developing a default prediction model in P2P lending that focuses on finding the class balancing method that has the best performance, which is depicted in Figure 1

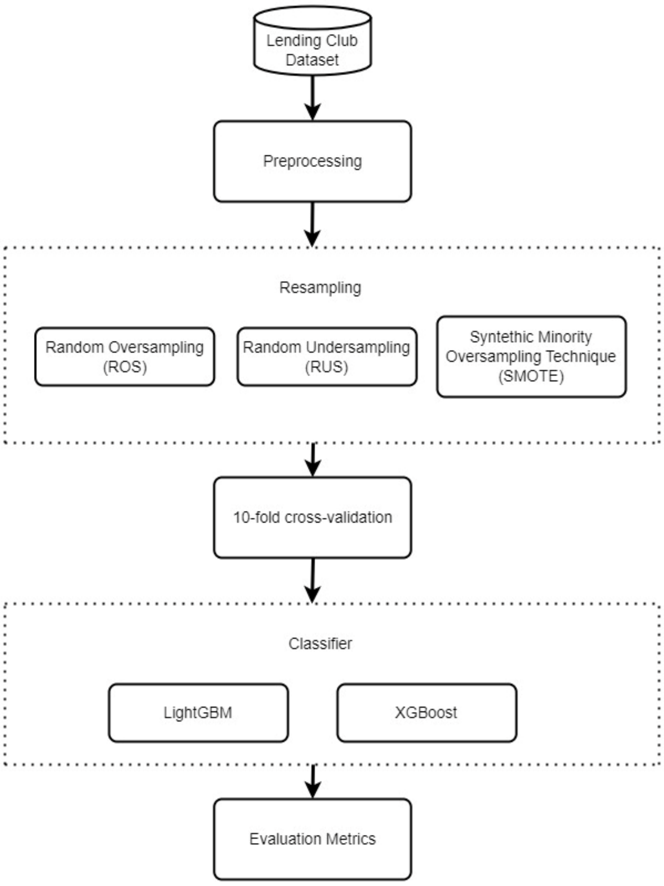

Figure 1. Proposed research framework

Lending club dataset

Lending club is a P2P lending platform that provides loan history data that can be accessed publicly via the Kaggle platform. This research uses a sample of lending club loan data from 2007-2019. The Lending Club dataset contains 396,029 row records, and 27 features, in detail can be accessed through the link https://www.kaggle.com/datasets/jeandedieunyandwi/lending-club-dataset.

Preprocessing

To prepare the data, two methods were carried out, namely cleaning missing values and changing the category type to numeric. The first step is to remove any features and events where more than 30% of the data is missing. The remaining missing values were then imputed using mode for categorical data or mean for numerical features. Then, the string data type is converted to a numeric type, and the categorical features are one-hot encoded. After the data preprocessing process, will see the imbalance ratio of the number of classes. If a data set has an imbalanced number of classes, it can be calculated with the imbalanced ratio (IR) defined as Eq. (1).

$$IR = \frac{N_{maj}}{N_{min}} \quad (1)$$

where N_{maj} is the number of major classes, and N_{min} is the number of minor classes. When $IR > 1$ means the dataset has imbalanced class sizes, whereas if $IR = 1$ then the class sizes are definitely balanced. In this study, it is important to pay attention to the IR value of the data that will be used as a sample to assess the performance of the data balancing model (DBT) for each classifier.

Data balancing model

By altering the distribution class, the resampling procedure balances the datasets. It is separated into two techniques. The first technique is undersampling, which causes the size of the majority class to approach that of the minority class. The second technique, known as over-sampling, uses the minority class to enlarge it to a size that is comparable to the majority class [29].

Undersampling

To ensure that the datasets are balanced, a subsample of the majority class is chosen whose size corresponds to the set of minority classes. However, because it discards certain crucial information, it could create another problem. Random undersampling is another sort of undersampling that randomly omits data from the majority class until the class distribution is balanced. We used Random Undersampling (RUS) as an undersampling technique in this study. Undersampling techniques often act in two ways, by removing noisy instances, or simply reducing instances using heuristics or even randomly.

Oversampling

To balance the distribution of the dataset, the oversampling approach generates additional data on minority groups. By randomly duplicating the data, the random oversampling (ROS) approach is an easy way to increase minority class data. Another methodology for oversampling is called SMOTE [30], or synthetic minority oversampling technique. Consider a feature vector x_i of minority examples, where m is the minority example's closest neighbor in the feature space. Then, once the distribution is balanced, fresh data for the minority class are produced by interpolating between m and x_i. The new SMOTE oversampling technique known as "Borderline SMOTE" solely oversamples minority class borderline data [31]. If the number of x_i nearest neighbor that belongs to the majority class which fit $\frac{m}{2} < \left|x_j \cap majority\right| < m$, define the x_i near the borderline and form the new data.

Cross-validation

In this work, our model was trained, validated, and tested using k-fold cross-validation. In the k-fold cross-validation approach, k- subsets are randomly selected from the original data set. In one of the k rounds, each of the k subsets serves as a test data set. For model training and model fitting, the remaining *k-1* subsets are employed. This strategy, according to [32], lessens the effect of data reliance. To put it another way, the classifier is graded sequentially on the whole data set, which reduces the chance that the performance of a classifier depends on the testing set selection. Additionally, [33] emphasizes that using k-fold cross-validation is a guarantee of the validity of the results. In our paper, we specifically use 10-fold cross-validation.

Classifier

In this section, we will explain the machine learning models in this study. Resampling as a class balancing is used for the LightGBM and XGBoost classifiers.

LightGBM

According to [34] LightGBM is a distributed gradient boosting framework first developed by Microsoft for machine learning algorithms. LightGBM is an extension of GBDT algorithms such as XGBoost. Ke et al. suggested two methods, gradient-based one-sided sampling (GOSS) and exclusive feature bundling (EFB), to reduce the amount of data and features without compromising the predictive ability of the model. These methods can help achieve a balance between accuracy and efficiency.

GOSS is an algorithm that maximizes data reduction while preserving accuracy. The gradient may be utilized as a gauge of sample weights since the approach predicts that samples with large gradients would significantly affect the gain criterion. By randomly selecting samples, the GOSS approach preserves all samples with big gradients and discards some samples with minor gradients. The gain is then calculated after weighting the sampled tiny gradient data. As a result, the algorithm will concentrate more on cases with insufficient training and will not significantly alter the distribution of the initial data.

Another method that maximizes fewer features while retaining accuracy is EFB. The EFB method can convert many mutually exclusive features into higher density features. It avoids unnecessary counting of zero-valued features because the high-dimensional feature space is quite sparse and many features are almost mutually exclusive to each other. In this work, the optimal set of hyperparameters was discovered using both techniques.

XGBoost

The Gradient Boosting Decision Tree (GBDT), which was developed by [35] is extended in XGBoost, which was proposed by [36]. Although the GBDT uses ensemble learning, it learns in little steps. Each tree that is created using this approach is built upon its predecessor. This implies that the original design is always used, but that a new feature is always introduced to make up for the faults of the prior design. The XGBoost models utilize a more effective approximation approach (also known as the histogram algorithm) in place of the greedy algorithm used in GBDTs, which not only increases computing efficiency but also lessens the overfitting issue.

Additionally, we may implement out-of-core computing, parallel computing, and decentralized computing in this system, which decreases training time and effectively handles massive amounts of data. Please refer to Chen and Guestrin [36] for more information since this work focuses more on the empirical investigation and development of a credit-scoring model.

Evaluation metrics

A confusion matrix is a useful tool for classifier assessment in machine learning. The predicted label of a class is represented by each column of the matrix, while the actual label of a class is represented by each row. The confusion matrix design for binary classification is depicted in Table 1.

Table 1. Confusion matrix

	Predicted Value	
Actual Value	**Negative**	**Positive**
Negative	True Negative (TN)	False Positive (FP)
Positive	False Negative (FN)	True Positive (TP)

False positive (FP) and false negative (FN) are terms for classifiers that predict correctness and incorrectness. Furthermore, accuracy is the percentage of predictions correctly predicted by our model. Since its representation predicts the positive rate across all true positive data, the sensitivity representation is referred to as the precision rate. We use the F1 score, which is the harmonic mean of recall and precision, to combine the both. All evaluation metrics can be shown in Eq. (2).

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

$$Recall = \frac{\mathrm{TP}}{(\mathrm{TP}+\mathrm{FN})} \tag{2}$$

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

$$\text{F1} - \text{Score} = \frac{2Precision \times Recall}{Precision + Recall}$$

RESULTS AND DISCUSSIONS

In preprocessing, features were eliminated with more than 30% missing values, and excessive unique values resulting in 24 features and 395,219 records. Then, of the 24 features there are 7 features that are category type, this means that the data cannot be processed into the classification modeling stage. To handle this, the data type is converted to numeric using one-hot encoding, in detail can be seen in Table 2.

Table 2. Converting data type using one-hot encoding

Categorical Feature	Feature Number After One-Hot Encoding
sub_grade	35 features
verification_status	3 features
purpose	7 features
initial_list_status	1 feature
application_type	2 features
home_ownership	5 features
zip_code	9 features

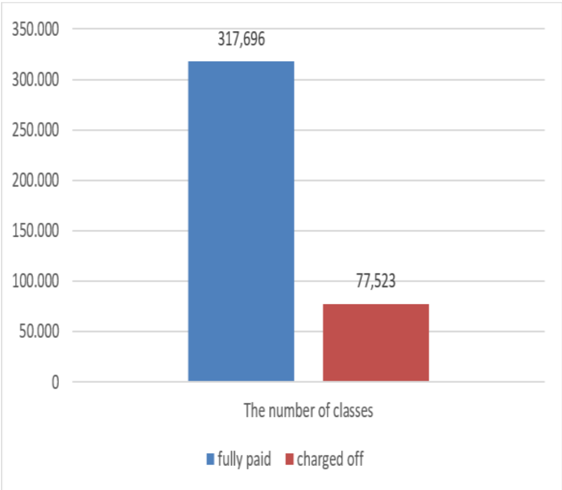

Figure 2. The proportion number of loan status classes (IR=4.098)

Data visualization results were also obtained, which are presented in Figure 2, which are the number of 'loan status' classes with a value of 0 as 'charged off' (77,523) records and 1 as 'fully paid' (317,696) records. Based on Eq. (1) then the imbalanced ratio (IR) value is 4,098, which means IR>1, so the class size is said to be imbalanced.

The performance measurement results of the impact of using the data balancing model can be seen from two classifiers, which are LightGBM and XGBoost. Each classifier has been combined with a data balancing model, namely random oversampling (ROS), random underside (RUS), and synthetic minority oversampling technique (SMOTE), forming the LightGBM model as follows: LightGBM+ROS, LightGBM+RUS, and LightGBM +SMOTE. Likewise, for XGBoost it is as follows: XGBoost+ROS, XGBoost+RUS, and XGBoost+SMOTE. How the model performs is validated with 10-fold cross-validation and evaluated based on accuracy, recall, precision, and F1-score.

The difference in the amount of data from different undersampling and oversampling techniques. In the undersampling technique, the amount of data in the major class is trimmed so that it is equal to the number of minor classes. Conversely, in the oversampling technique, the number of minor classes will increase so that it is equal to the number of major classes.

Table 3. The amount of data in the class is based on the data balancing model

Data balancing model	Fully paid	Charged off	Total
Original	317,696	77,523	395,219
Undersampling	77,523	77,523	155,046
Oversampling	317,696	317,696	635,392

The experimental results show the value of measuring the performance of the data balancing model on two classifiers, namely LightGBM and XGBoost. Classifier performance is seen from the level of accuracy, recall, precision and F1-score. In LightGBM, it was tested using the original data size with each data balancing model including ROS, RUS, and SMOTE showing significant performance results. The performance results for each data balancing model against the LightGBM classifier can be seen in Table 4. We can see the results that have been obtained from the performance of the data balancing model on the LightGBM classifier, from the results that have been achieved the best performance was obtained by the LightGBM+SMOTE model with an accuracy of 92.56%, recall of 99.01%, precision of 87.70%, and FI-score is 93.01%. Then, in Table 5 presents the results of the performance of each data balancing model on the XGBoost classifier. In the XGBoost classifier, the performance of the data balancing model is not much different from LightGBM, where we can see that F1-score 92.76%.

Table 4. Performance comparison of data balancing model on LightGBM classifier

Data Balancing Model	Accuracy	Recall	Precision	F1-Score
LightGBM+Imbalanced data	88.95%	**99.30%**	**88.39%**	**93.53%**
LightGBM+ROS	81.04%	80.64%	81.29%	93.48%
LightGBM+RUS	80.55%	80.43%	80.62%	80.53%
LightGBM+SMOTE	**92.56%**	99.01%	87.70%	93.01%

Table 5. Performance comparison of data balancing model on XGBoost classifier

Data Balancing Model	Accuracy	Recall	Precision	F1-Score
XGBoost+Imbalanced data	88.92%	**98.93%**	**88.60%**	93.48%
XGBoost+ROS	82.04%	81.28%	82.54%	81.90%
XGBoost+RUS	80.39%	80.10%	80.57%	80.33%
XGBoost+SMOTE	**92.32%**	98.41%	87.72%	**92.76%**

After looking at the performance results of the data balancing models on LightGBM and XGBoost, it can be discussed that the data balancing model with the best performance is SMOTE, with accuracy and F1 scores that are significantly superior to ROS and RUS. Meanwhile, the lowest performance was shown by RUS for each classifier. This shows that SMOTE with its performance that increases the number of samples based on nearest neighbor minority examples or in feature space is recommended to overcome class imbalance in the Lending Club dataset. In this case, we find that the larger the data size used as a model training sample, the superior performance obtained by the classification model in predicting credit risk in P2P lending.

We compare the performance of the model in this study with the results of previous research that has used P2P lending data sets. Table 6 presents the proposed comparative study with previous related works. It can be seen that the model experimented in this study generally outperforms all models in previous research. In the overall summary of the results of previous research, SMOTE has superior performance as a data balancing model, for each classifier including LR, DT, RF, MLP, and NN.

Table 6. Comparison of result with previous related works

Study	Classifier	Data Balancing Model	Accuracy	F1-Score
[37]	LR	SMOTE	73.06%	-
	MLP		69.08%	-
	RF		71.61%	-
[38]	LR	SMOTE	-	88.78%
	DT		-	84.31%
[39]	RF	SMOTE	92.00%	-
	NN		87.56%	-
	LR		87.48%	-
This study	LightGBM	SMOTE	92.56%	93.01%
	XGBoost		92.32%	92.76%

This research provides an important contribution to existing research regarding credit risk prediction in P2P lending services. The research results show that the LightGBM+SMOTE and XGBoost+SMOTE models have a high level of performance, with an accuracy of 92.56% and 92.32%, respectively. Comparing the model performance results in this study with the model performance in previous studies, LightGBM+SMOTE and XGBoost+SMOTE significantly outperformed the models in previous studies. For example, the SMOTE data balancing model in this study has better performance than SMOTE which also has superiority in previous studies on classifiers such as LR, DT, RF, MLP, and NN. This shows that the data balancing model used in this research, which are the combination of SMOTE with LightGBM and XGBoost, has brought significant improvement in results in predicting credit risk in P2P loans. This is because LightGBM and XGBoost are tree-based classifiers that have the advantage of speed in the classification process, of course they are effective as classifiers with large sample data. Recalling the idea from the research of Chawla et al. that the majority sample data or in this case those with larger data sizes will have the opportunity to create synthetic neighbours that can improve SMOTE performance [27]. These results provide an important contribution as a basis for further research in research in this domain.

CONCLUSION

Credit risk prediction in P2P lending is a popular and growing research to find the best solution regarding the performance of machine learning models. However, P2P lending has fundamental problems because it has imbalanced data classes, where this problem can reduce the performance of the classification model. In this study, we have successfully used machine learning algorithms to predict credit risk in P2P lending and used data balancing models such as ROS, RUS, and SMOTE to process imbalanced data sets. To evaluate the experiments carried out, this research used the Lending Club dataset with IR = 4,098. In the experimental results, it was found that SMOTE as a data balancing model obtained the best performance compared to other data balancing models. The model performance was validated using 10-fold cross-validation, then the accuracy level produced superior performance results on LightGBM+SMOTE with an accuracy of 92.56%, followed by XGBoost+SMOTE with an accuracy not much different, which is 92.32%. Meanwhile, the research has experimental limitations which only test three data balancing models and two classifiers. So, as a recommendation for future research, the researcher can test more data balancing models other than this study, with more varied classifiers.

REFERENCES

[1] I. Lee and Y. J. Shin, “Fintech: Ecosystem, business models, investment decisions, and challenges,” *Bus Horiz*, vol. 61, no. 1, pp. 35–46, 2018, doi: 10.1016/j.bushor.2017.09.003.

[2] Z. Fang, J. Zhang, and F. Zhiyuan, “Study on P2P E-finance platform system: A case in China,” *Proceedings - 11th IEEE International Conference on E-Business Engineering, ICEBE 2014 - Including 10th Workshop on Service-Oriented Applications, Integration and Collaboration, SOAIC 2014 and 1st Workshop on E-Commerce Engineering, ECE 2014*, pp. 331–337, 2014, doi: 10.1109/ICEBE.2014.64.

[3] T. Balyuk, “FinTech Lending and Bank Credit Access for Consumers ∗,” *Rotman School of Management Working Paper No. 2802220*, 2021.

[4] 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.

[5] Q. Zhu, W. Ding, M. Xiang, M. Hu, and N. Zhang, “Loan Default Prediction Based on Convolutional Neural Network and LightGBM,” *International Journal of Data Warehousing and Mining*, vol. 19, no. 1, pp. 1–16, 2022, doi: 10.4018/IJDWM.315823.

[6] W. Hou, X. Wang, H. Zhang, J. Wang, and L. Li, “A novel dynamic ensemble selection classifier for an imbalanced data set: An application for credit risk assessment,” *Knowl Based Syst*, vol. 208, p. 106462, 2020.

[7] L. E. B. Ferreira, J. P. Barddal, H. M. Gomes, and F. Enembreck, “Improving credit risk prediction in online peer-to-peer (p2p) lending using imbalanced learning techniques,” in *2017 IEEE 29th International Conference on Tools with Artificial Intelligence (ICTAI)*, IEEE, 2017, pp. 175–181.

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

[9] P. Pampouktsi *et al.*, “Techniques of Applied Machine Learning Being Utilized for the Purpose of Selecting and Placing Human Resources within the Public Sector,” *Journal of Information System Exploration and Research*, vol. 1, no. 1, pp. 1–16, 2023.

[10] M. A. Muslim, Y. Dasril, H. Javed, W. F. Abror, D. A. A. Pertiwi, and T. Mustaqim, "An Ensemble Stacking Algorithm to Improve Model Accuracy in Bankruptcy Prediction," *Journal of Data Science and Intelligent Systems*, vol. 1, no. 1, 2023.

[11] P. Kaur and A. Gosain, "Comparing the behavior of oversampling and undersampling approach of class imbalance learning by combining class imbalance problem with noise," in *Advances in Intelligent Systems and Computing*, 2018. doi: 10.1007/978-981-10-6602-3_3.

[12] X. Yang, W. Fan, L. Wang, S. Yang, and W. Wang, "Risk control of online P2P lending in China based on health investment," *Ekoloji*, vol. 28, no. 107, pp. 2013–2022, 2019.

[13] C. P. Swee, F. Meziane, J. Labadin, I. Technology, and K. Samarahan, "Credit Risk Prediction for Peer-To-Peer Lending Platforms : An Explainable Machine Learning Approach," *Journal of Computing and Social Informatics*, vol. 1, no. 2, pp. 1–16, 2022.

[14] C. Rao, M. Liu, M. Goh, and J. Wen, "2-stage modified random forest model for credit risk assessment of P2P network lending to 'Three Rurals' borrowers," *Applied Soft Computing Journal*, vol. 95, 2020, doi: 10.1016/j.asoc.2020.106570.

[15] P. Teply and M. Polena, "Best classification algorithms in peer-to-peer lending," *The North American Journal of Economics and Finance*, vol. 51, p. 100904, 2020.

[16] L. Zhu, "ScienceDirect A study study on on predicting predicting loan loan default default based based on on the the random random forest forest algorithm algorithm," *Procedia Comput Sci*, vol. 162, no. Itqm 2019, pp. 503–513, 2020, doi: 10.1016/j.procs.2019.12.017.

[17] Y. Chen and Z. Lin, "Business Intelligence Capabilities and Firm Performance: A Study in China," *Int J Inf Manage*, vol. 57, no. xxxx, p. 102232, 2021, doi: 10.1016/j.ijinfomgt.2020.102232.

[18] R. Ge, J. Feng, B. Gu, and P. Zhang, "Predicting and Deterring Default with Social Media Information in Peer-to-Peer Lending Predicting and Deterring Default with Social Media Information in Peer-to-Peer Lending," *Journal of Management Information Systems*, vol. 34, no. 2, pp. 401–424, 2017, doi: 10.1080/07421222.2017.1334472.

[19] W. Bao, N. Lianju, and K. Yue, "Integration of unsupervised and supervised machine learning algorithms for credit risk assessment," *Expert Syst Appl*, vol. 128, pp. 301–315, 2019, doi: 10.1016/j.eswa.2019.02.033.

[20] A. E. Omarini, "Peer-to-peer lending: business model analysis and the platform dilemma," *International Journal of Finance, Economics and Trade (*, vol. 2, no. 3, pp. 31–41, 2018.

[21] A. Bachmann *et al.*, "Online peer-to-peer lending-a literature review," *Journal of Internet Banking and Commerce*, vol. 16, no. 2, p. 1, 2011.

[22] C. Serrano-Cinca, B. Gutiérrez-Nieto, and L. López-Palacios, "Determinants of default in P2P lending," *PLoS One*, vol. 10, no. 10, pp. 1–22, 2015, doi: 10.1371/journal.pone.0139427.

[23] S. Cai and J. Zhang, "Exploration of credit risk of P2P platform based on data mining technology," *J Comput Appl Math*, vol. 372, p. 112718, 2020.

[24] L. Zhu, D. Qiu, D. Ergu, C. Ying, and K. Liu, "A study on predicting loan default based on the random forest algorithm," *Procedia Comput Sci*, vol. 162, pp. 503–513, 2019, doi: 10.1016/j.procs.2019.12.017.

[25] L. Vinod Kumar, S. Natarajan, S. Keerthana, K. M. Chinmayi, and N. Lakshmi, "Credit Risk Analysis in Peer-to-Peer Lending System," *2016 IEEE International Conference on Knowledge Engineering and Applications, ICKEA 2016*, no. July, pp. 193–196, 2016, doi: 10.1109/ICKEA.2016.7803017.

[26] N. Rout, D. Mishra, and M. K. Mallick, "Handling imbalanced data: a survey," in *International Proceedings on Advances in Soft Computing, Intelligent Systems and Applications: ASISA 2016*, Springer, 2018, pp. 431–443.

[27] N. V Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: synthetic minority over-sampling technique," *Journal of artificial intelligence research*, vol. 16, pp. 321–357, 2002.

[28] G. Chawla, L. R. Forest Jr, and S. D. Aguais, "Point-in-time loss-given default rates and exposures at default models for IFRS 9/CECL and stress testing," *Journal of Risk Management in Financial Institutions*, vol. 9, no. 3, pp. 249–263, 2016.

[29] C.-R. Wang and X.-H. Shao, "An improving majority weighted minority oversampling technique for imbalanced classification problem," *IEEE Access*, vol. 9, pp. 5069–5082, 2020.

[30] N. V Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: synthetic minority over-sampling technique," *Journal of artificial intelligence research*, vol. 16, pp. 321–357, 2002.

[31] H. Han, W.-Y. Wang, and B.-H. Mao, "Borderline-SMOTE: a new over-sampling method in imbalancedA data sets learning," in *International conference on intelligent computing*, Springer, 2005, pp. 878–887.

[32] S. L. Salzberg, “On comparing classifiers: Pitfalls to avoid and a recommended approach,” *Data Min Knowl Discov*, vol. 1, pp. 317–328, 1997.

[33] L. Huang and D. Chiang, “Forest rescoring: Faster decoding with integrated language models,” in *Proceedings of the 45th annual meeting of the association of computational linguistics*, 2007, pp. 144–151.

[34] G. Ke *et al.*, “LightGBM: A highly efficient gradient boosting decision tree,” *Adv Neural Inf Process Syst*, no. December, pp. 3147–3155, 2017.

[35] J. H. Friedman, “Greedy function approximation: A gradient boosting machine,” *Ann Stat*, vol. 29, no. 5, pp. 1189–1232, 2001, doi: 10.1214/aos/1013203451.

[36] T. Chen and C. Guestrin, “Xgboost: A scalable tree boosting system,” in *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, 2016, pp. 785–794.

[37] H. Qian, S. Zhang, B. Wang, L. Peng, S. Gao, and Y. Song, “A comparative study on machine learning models combining with outlier detection and balanced sampling methods for credit scoring,” *Expert Syst Appl*, 2021, [Online]. Available: http://arxiv.org/abs/2112.13196

[38] Y. Wang and X. Sherry Ni, “Predicting Class-Imbalanced Business Risk Using Resampling, Regularization, and Model Emsembling Algorithms,” *International Journal of Managing Information Technology*, vol. 11, no. 01, pp. 01–15, 2019, doi: 10.5121/ijmit.2019.11101.

[39] A. A. Hussin Adam Khatir and M. Bee, “Machine Learning Models and Data-Balancing Techniques for Credit Scoring: What Is the Best Combination?,” *Risks*, vol. 10, no. 9, 2022, doi: 10.3390/risks10090169.