Intelligent Systems with Applications 18 (2023) 200204

Contents lists available at ScienceDirect

Intelligent Systems with Applications

journal homepage: www.journals.elsevier.com/intelligent-systems-with-applications

New model combination meta-learner to improve accuracy prediction P2P lending with stacking ensemble learning*

Much Aziz Muslim [a,b,*], Tiara Lailatul Nikmah [b], Dwika Ananda Agustina Pertiwi [b], Subhan [b], Jumanto [b], Yosza Dasril [a], Iswanto [c]

[a] Faculty of Technology Management, Universiti Tun Hussein Onn Malaysia, Johor 86400, Malaysia
[b] Department of Computer Science, Universitas Negeri Semarang, Semarang 50229, Indonesia
[c] Department of Electrical Engineering, Universitas Muhammadiyah Yogyakarta, Bantul 55183, Indonesia

ARTICLE INFO

Keywords:
LightGBM
P2P lending
Default risk prediction
Stacking ensemble learning
Improve accuracy prediction

ABSTRACT

Peer-to-peer (P2P) Lending is a type of financial innovation that offers loans without intermediaries to individuals and companies. In the P2P lending system, there is a risk of default on the loan which causes the company to lose. Many studies have to reduce the risk of default by developing a classification model of prediction of default that focuses on increasing accuracy. However, the big problem with prediction is data imbalance and low performance classification algorithms. The purpose of this study is to improve the accuracy of default risk prediction by balancing the data and combining the stacking model ensemble with the meta-learner. The proposed new model consists of 3 optimization parts, the first is Synthetic Minority Oversampling Technique (SMOTE), the second is the selection of features and the third is stacking ensemble learning. The SMOTE method is used to balance the data, the feature selection LightGBM and stacking ensemble learning (LGBFS-StackingXGBoost) to optimize machine learning accuracy. A new model of stacking ensemble learning by combining three base-learner algorithms namely KNN, SVM and Random Forest into the XGBoost meta-learner algorithm. The model was tested using two datasets, namely the online P2P lending dataset and the lending club loan data analysis dataset. The evaluation results show that LGBFS-StackingXGBoost is the best model for both datasets. In the online P2P lending dataset, it received an accuracy of 99,982% and in the lending club loan data analysis dataset, it received an accuracy of 91,434%. This study shows that the accuracy of the prediction model can be improved using the LGBFS-StackingXGBoost method.

1. Introduction

Financial technology (fintech) is an example of innovation in the financial sector by utilizing information technology. One type of fintech that has gained popularity in the 20th century is a peer-to-peer lending (Perez et al., 2020). P2P lending is an online platform for peer-to-peer lending that functions as an intermediary between individual and business lenders and borrowers by providing microcredit services (Mota et al., 2018). Such loans allow borrowers to obtain microloans through online channels without the involvement of financial institutions like banks and credit card firms so it is easier, more transparent and faster than traditional banks (Ma et al., 2018).

P2P lending has grown quickly over the past decade and the loan transaction capacity has reached RMB 982 billion Yuan in 2015 (Yoon et al., 2019). Start-ups and expanding companies most frequently use this method to finance their operations as an alternative source of funds. Based on data from the LendingClub platform, the loan amount has reached US$13.4 billion by the end of 2015 (Zhao et al., 2018) and $12, 290 billion in US dollars by the end of 2019 (Perez et al., 2020).

The P2P lending platform has the concept of direct money lending without intermediaries. Therefore, lenders have greater freedom to select the preferred risk portfolio (Byanjankar et al., 2021). Because failures often cause huge losses for lenders and can threaten the development of the P2P lending system. Therefore, a strategy is needed to reduce the existence of loan default risk, and find effective ways to manage risks that may occur (Luo & Yan, 2022).

* Corresponding author.
E-mail addresses: a212muslim@mail.unnes.ac.id (M.A. Muslim), tiaralaila21@gmail.com (T.L. Nikmah), dwikapertiwi13@gmail.com (D.A.A. Pertiwi), subhan@mail.unnes.ac.id (Subhan), jumanto@mail.unnes.ac.id (Jumanto), yosza@uthm.edu.my (Y. Dasril), iswanto_te@umy.ac.id (Iswanto).

https://doi.org/10.1016/j.iswa.2023.200204
Received 26 December 2022; Received in revised form 9 February 2023; Accepted 10 February 2023
Available online 11 February 2023

Therefore, many researchers develop techniques to reduce the risk of default by data mining. Data mining is a machine learning method that aims to search for information. The use of data mining usually aims to get the highest accuracy. A wide variety of optimizations with various algorithms have been carried out in pursuit of very high accuracy. The classification method is a type of data mining technique for making predictions on a class of data (Ashari et al., 2016; Budiman & Ifriza, 2021; Damayanti et al., 2022; Falasari & Muslim, 2022; Hazimah et al., 2022; Muslim et al., 2018; Prasetiyo et al., 2020).

Related research has applied a classification method for predicting default by carrying out a combination of preprocessing techniques to overcome unbalanced data (Jadwal et al., 2022; Muslim et al., 2021; Papouskova & Hajek, 2019; Prasetiyo et al., 2021b) and experimenting with various feature selection methods (Mardiansyah et al., 2022), (Prasetiyo et al., 2021a). One of the most popular classification models today is the stacking ensemble classifier. Stacking ensemble uses the concept of meta-learners to find the best results by combining classifiers from several base-learner algorithms (Ragab et al., 2021). However, we found that research by Kun et al. (Kun et al., 2020), still uses the default meta-learner model for the final estimator (Kang et al., 2015; Liang et al., 2022; Xia et al., 2021). Furthermore, an effective improved meta-learner model is the contribution of this research. We propose a boost-based meta-learner model that is XGBoost as the final estimator, and it succed in improving accuracy performance.

Our contributions research is: i) improve the accuracy model of loan default prediction on the Lending Club platform using the stacking ensemble with the XGBoost meta-learner, ii) optimized by adding LightGBM feature selection and iii) the SMOTE oversampling method.

The remainder of this paper is organized as follows. In Section 2, provides a literature review of the study that has been conducted on loan default prediction in P2P lending and the improvement of proposed method. In Section 3, the framework of SMOTE-oversampling, LGBFS, stacking ensemble, evaluation model, and we present a description of the dataset. Then, comparison of results of this experiments in Section 4. Finally, we conclude the paper by summarizing our contributions and discussing future research directions in Section 5.

2. Literature Review

Various approaches to predicting credit risk are categorized into three developmental stages: competency, statistics and machine learning based assessment forecasting. At first, skill-based methods were mostly used to evaluate credit risk. Subjective judgments serve as the basis for the appraisal process, making judgments a product of the rater's expertise. Song et al. (2020) confirmed that it is difficult to even for experienced borrowers to make reasonable decisions in many situations. As data science advances, statistics becomes more effective and efficient model-based techniques for credit risk assessment of loans were developed. Some statistical technique-based research, such as LightGBM (Ma et al., 2018; Zhou et al., 2019), XGBoost (Wang et al., 2022), (Li et al., 2020), K-NN (Sagar et al., 2022), SVM (Wang & Li, 2019), (Liang & Cai, 2020), Random Forest (Zhu et al., 2019), (Li et al., 2021) and ensemble learning (Li et al., 2021).

This study used the same dataset as Muslim et al. (2022)) namely the Online P2P Lending dataset that uses the single classifier LightGBM and gets an accuracy of 95.64%. Then in Ma et al. (2018), the authors propose an analysis of the same dataset we examine in our paper, i.e., of the Lending Club dataset (Hou et al., 2020; Ruyu et al., 2019).

In another study, there was a model development carried out by Xia et al. (2021) by improvising heterogeneous stacking ensemble to improve Los Given Default (LGD) forecasting to predict default on P2P lending platforms based on the LendingClub dataset wich gets the best mean square error value of 0.044. Wang et al. (2017) used Random Forest (RF) to estimate the default probability and exact default period of the loan applicant. Random Forest outperforms other models, such as Logistic regression, the Cox proportional hazards model, and the conventional mixing model, effectively reduce the error rate in assessing the value of defaulting customers as non-defaulting customers.

On the other hand, research conducted by Kun et al. (2020) using stacking algorithms in his research resulted in a high accuracy of 98.16%. However, Kun's research only uses the default meta learner on its stacking model. Therefore, this study tried to use a new meta learner model in the stacking method to get high accuracy from previous studies.

In addition, some of the stacking meta-learners used in previous studies include Logistic Regression (Song et al., 2020), Naïve Bayes (Xiong et al., 2021) dan Random Forest (Khochare et al., 2020). Then, the stacking ensemble model can be improved by adding feature selection methods. One of them is by using the LightGBM algorithm as done by Li et al. (2021), Zhang et al. (2022) and by handling data imbalances as done by Papouskova & Hajek (2019)), Muslim et al. (2022) and Chen et al. (2021). Summary of default loan prediction research which is shown in Table 1.

In the present paper, we propose a new model based on a stacking ensemble to predict loan default. We propose a synthetic minority oversampling technique (SMOTE) model (Mukherjee & Khushi, 2021) to balance class data, then to select the best features we implement the LightGBM feature selection (LGBFS) model (Zhang et al., 2022). In the classification problem where each loan is classified as default or non-default, we propose a stacking ensemble model (Kun et al., 2020) that we improve on base learner and meta learner, we combine three classifiers namely KNN (Sagar et al., 2022), SVM (Wang & Li, 2019), and random forest (Zhu et al., 2019) as base learner, for our meta learner propose a boost-based model is the XGBoost algorithm (Susan et al., 2021). From the series we propose, we call the model LGBFS-StackingXGBoost, we evaluate the proposed model by assessing how high the resulting accuracy is for predicting loan default on the P2P Lending platform.

3. Methodology

In this study, the P2P loan default risk analysis uses several stages, namely data preprocessing, data oversampling, feature selection and model evaluation. The research framework is shown in Fig. 1.

3.1. Data Description

The study's dataset originated from the Kaggle repository, namely the online P2P lending dataset at the following link: https://www.kaggle.com/datasets/skihikingkevin/online-p2p-lending. The dataset contains Lending Club loan data for the period 2012-2018 which contains

Table 1
Summary of default loan prediction research.

Paper	Aims	Summary
(Zhu et al., 2019)	Default loan prediction using Random Forest and SMOTE	The Random Forest algorithm outperforms the algorithm logistic regression, decision trees, and other machine learning algorithms in predicting sample defaults. SMOTE method can overcome the problem of data imbalance in the dataset.
(Kun et al., 2020)	Credit default predict uses a stacking ensemble.	The stacking model consists of several single classifiers models such as ANN (Artificial Neural Network), RF (Random forest), AdaBoost (Adaptive Boosting) and XGBoost (Gradient boosting) algorithms. Stacking has better overall performance than single classifiers
(Li et al., 2021)	Network lending risk prediction using the CNN-stacking model	The CNN-stacking model is better than the single classifiers model in risk prediction accuracy

Fig. 1. Prediction methods of P2P lending.

2,875,146 user loans with 18 attributes namely loan number, amount borrower, term, borrower rate, installment, grade, listing title, principal balance, next payment due date, principal paid, interest paid, late fees paid, debt sale proceeds received, last payment date, days past due, origination date, loan status description and data source.

The second dataset is the lending club loan data analysis dataset, which is also taken from the Kaggle repository at the following link: https://www.kaggle.com/datasets/urstrulyvikas/lending-club-loan-data-analysis. The dataset contains Lending Club data for 2007-2015 which contains 9578 user loans with 14 attributes, namely credit policy, purpose, int rate, installment, log annual inc, dti, fico, revol bal, revol util, delinq 2yrs, days with cr line, pub rec, not fully paid and inq last 6mths.

Fig. 2 shows the results of data exploration analysis to determine the correlation between numerical features in the Online P2P Lending dataset. And in Fig. 3, we know the results of data exploration analysis to find out the correlation between numerical features in the Lending Club Loan Data Analysis dataset.

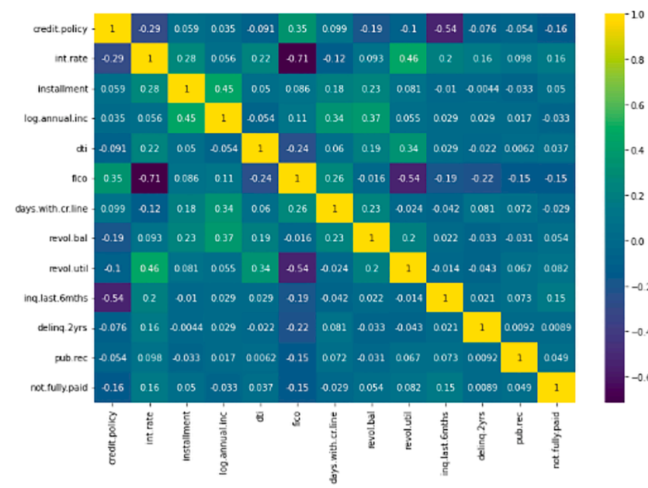

Fig. 3. Correlation matrix of lending club loan data analysis dataset.

3.2. Preprocessing Data

Data preprocessing is the process of cleaning data that can interfere with the prediction process (Mustaqim et al., 2020). Data that can interfere with the analysis process are missing values and outliers. Missing values are checked using a function from the Sklearn library and outliers are checked using the interquartile range (IQR) metric (Anagnostou et al., 2021). Columns with many missing values and rows with outliers will be deleted because they can disrupt model performance. Then do the encoding on non-numeric data. Non-numeric variables will be converted into numeric data by encoding techniques (Lopez-Arevalo et al., 2020). In the first dataset there is a non-numeric variable 'grade' which is encoded using the LabelEncoder function from the Sklearn library. And in the second dataset there are 'purpose' non-numeric variables which are encoded data using the one hot encoding method which uses the get_dummies function from the Pandas library.

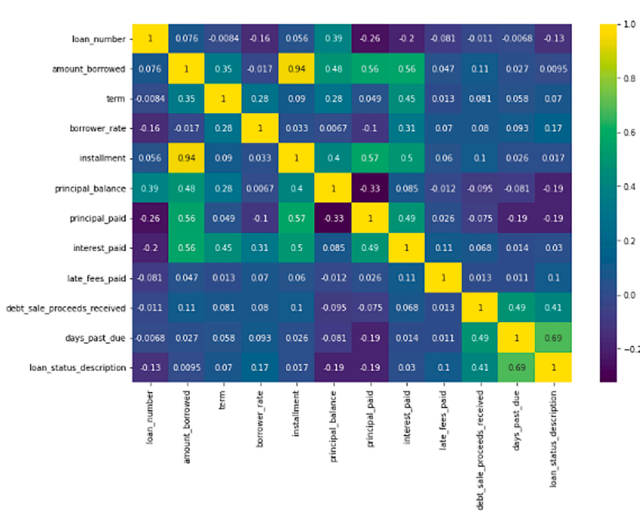

Fig. 2. Correlation matrix of online P2P lending dataset.

3.3. Oversampling Data

After the data preprocessing stage was carried out, there was a very large difference in the amount of data in the target variable 'loan_status' in the normal and default categories. This large difference can complicate the model in conducting learning.

In this study, researchers will handle class imbalances with resampling techniques (Mienye & Sun, 2021). The resampling technique used in this study uses the Oversampling method. The oversampling strategy was chosen because it increases the dataset that is lacking for minority classes and without reducing the amount of data (Lee & Kim, 2021). The oversampling used is the Synthetic Minority Oversampling Technique (SMOTE). Several previous studies have used SMOTE to address class imbalances in P2P loan credit risk prediction (Li et al., 2021), (Jadwal et al., 2022).

Synthetic data are used to equalize the number of instances in the minority class and the majority class. Synthetic data is added to the minor class dataset (in this case, class 1) using the SMOTE method (in this case class 0) (Wardoyo et al., 2022). In this method each sample of minority class is taken, and a synthetic sample is created by examining one or more neighbors of the k-sample (Pradipta et al., 2021). The results of the data oversampling process can be shown in Figs. 4 and 5.

3.4. LightGBM feature selection (LGBFS)

The feature selection process uses the LightGBM algorithm. By outlining the necessary features and learning the relationships between variables, feature selection helps in a better understanding of the data (Mohammadi et al., 2019). Feature selection is performed to reduce feature dimensions and lower computational complexity during the classifier training stage (Devan & Khare, 2020).

LightGBM uses histogram-based techniques and enhanced network connectivity to maximize parallel learning as well as accelerate training and use less memory (Machado et al., 2019). Feature selection with LightGBM determines feature weight values that are calculated by using the influence of features on predictive analysis results (Yang et al., 2021).

In this study using the feature selection technique with the feature_importance_ value from the LightGBM algorithm (Sarıkaya et al., 2022). Reweighting is done based on the misclassification generated by the selected features. The best features from each round, according to a feature rating, are added to the selected subset.

3.5. Stacking ensemble

The ensemble method is the merging of several sets of models to obtain a more accurate model (Abdar et al., 2020). The ensemble method can effectively reduce misclassification and is able to improve the performance effectiveness of a single classification model (Vianita et al., 2021). The ensemble method combines different sets of models to create a more accurate model (Ragab et al., 2021). The ensemble method has the power to significantly lower misclassification and increase the efficiency of a single classifier (Mohammed & Kora, 2022).

Stacking is a combination of several models of different types using the concept of a meta-learner (Shorfuzzaman, 2021). The first training data is used to train base learners at the first layer to generate a new dataset that is reused to train the second layer meta-learner. The first learner output is an input feature while the original label is still considered the new training data label. First-level base learners use several different classification algorithms (Xiong et al., 2018). The use of a meta-learner makes the stacking model stronger in its classification accuracy (Chaudhary et al., 2017), (Gupta et al., 2020).

On this ensemble stacking using the XGBoost meta-learner. XGBoost is a development of the gradient boost algorithm of several additional processes so that it is more powerful. Those additional processes are trimming, newton boosting, and extra randomization parameters (Jha et al., 2022). The process of pruning or proportional shrinkage of leaf nodes is utilized to increase the model's generalization. The newton boosting process is the process of providing a direct route so that it does not require a gradient drop. The parameter randomization process aims to reduce the correlation between trees to increase the strength of the ensemble algorithm.

3.6. Evaluation model

Confusion Matrix is able to present predictions and the actual state of the data generated by machine learning algorithms. In the Confusion Matrix, True Positives(TP), True Negatives(TN), False Positives(FP) and False Negatives(FN) are the correct number of positive classes, the number of false positive classes, correct negative class and incorrect negative class on the data (Al-Asadi & Tasdemír, 2021). By using this matrix, Accuracy, Recall and f1 score can be determined. Accuracy is determined by the proportion of samples that were correctly assigned to the overall sample size for the test dataset (Chicco & Jurman, 2020) as shown in Equation (2).

$$Accuracy = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{TP} + \mathrm{TN} + \mathrm{FP} + \mathrm{FN}} \tag{2}$$

Recall is a ratio measure of true positive predictions to all true positive data, namely TP and FN. Eq. (3) is the formula for calculating Recall.

$$Recall = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \tag{3}$$

The F1 score, or the average recall value (the ability to find all positive samples of data) and precision (the ability not to label negative classes as positive classes). The F1 score takes values in the range [0, 1]. Eq. (4) is the formula for calculating the f1 score.

$$f1\ score = \frac{2\ \mathrm{x\ precision\ x\ recall}}{\mathrm{precision} + \mathrm{recall}} \tag{4}$$

4. Result and discussion

This research was tested with 2 Lending Club datasets, namely the Online P2P Lending dataset which has 18 attributes and the Lending Club Loan Data Analysis dataset which has 14 attributes. This dataset is still raw data that has many missing values and outliers, so it must go through the data preprocessing stage first. The preprocessing stage consists of cleaning data and encoding non-numeric data. Data cleaning aims to remove missing values or missing data and outlier data. So from 2,875,146 data in the Online P2P Lending dataset it produces 2,120,012

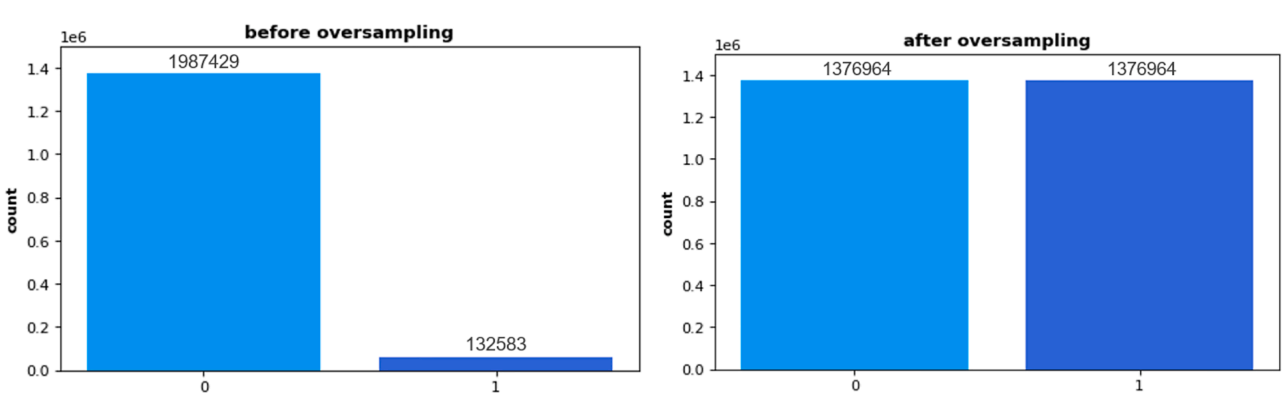

Fig. 4. Results of oversampling from online P2P lending.

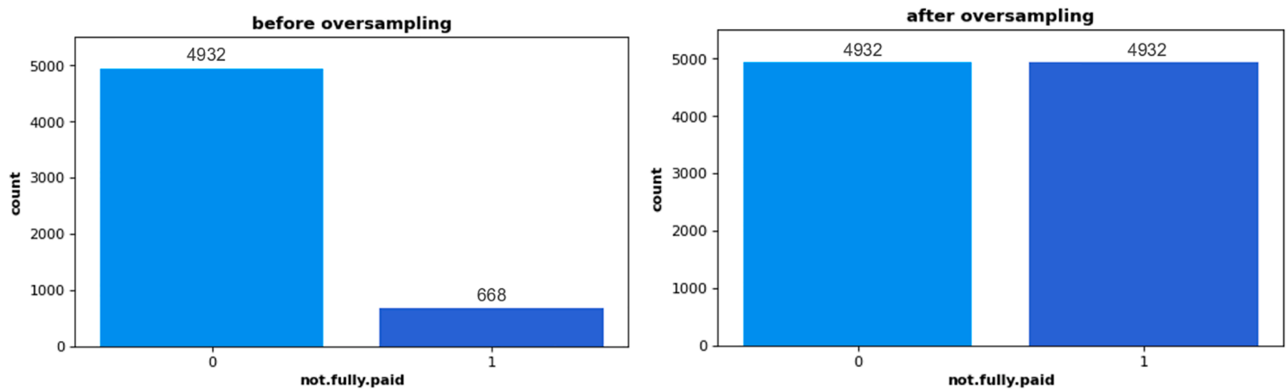

Fig. 5. Results of oversampling from lending club loan data analysis.

clean data and from 9578 data in the Lending Club Loan Data Analysis dataset it produces 5600 clean data. Then data encoding is performed to convert non-numeric data into numeric data, namely the 'grade' and 'purpose' attributes.

After that, the clean data goes through the oversampling stage with SMOTE which balances the number of classes 0 and 1. Then the data enters the LightGBM feature selection stage. At the feature selection stage, the 10 best features of the Online P2P Lending dataset were selected and the 9 best features of the Lending Club Loan Data Analysis dataset were selected to be used for model learning. The feature selection process uses the feature_importance_ attribute from the Sklearn library in the LightGBM algorithm. The results of selecting this feature show the value of each feature's weight. The feature that has the highest weight and weight value is the feature that has the strongest correlation with the predicted results. The results of ranking important features in the Online P2P Lending dataset are displayed in Fig. 6.

After testing for each number of features, 10 features obtained the highest accuracy. The test results for each number of features are shown in the graph in Fig. 7.

While the results of ranking important features in the dataset for Lending Club Loan Data Analysis are shown in Fig. 8.

After testing for each number of features, 9 features obtained the highest accuracy. The test results for each number of features are shown in the graph in Fig. 9.

Then the sampling data will go into the model. This study uses models from four single classification algorithms namely KNN, SVM and Random Forest, then the ensemble stacking method from combining the four algorithms and the stacking model using the meta-learners XGBoost, LightGBM, and AdaBoost. The results obtained to state that the results of the combination of stacking and among others, the XGBoost meta-learner produces the highest results. The test results were evaluated for the model using the Confusion Matrix. Evaluation results

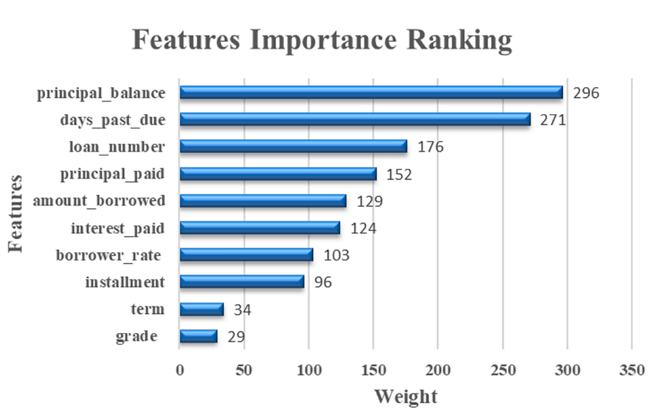

Fig. 6. Plot feature importance from online P2P lending data.

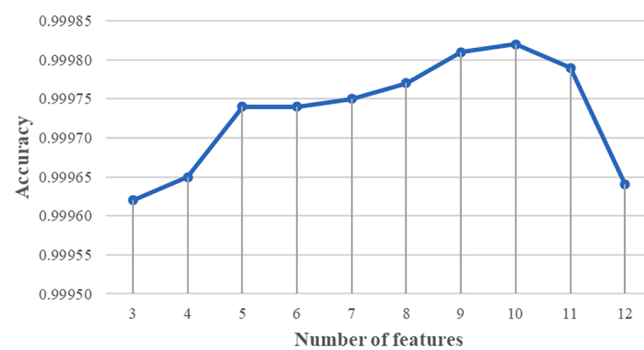

Fig. 7. Graph of feature trials on online P2P lending data.

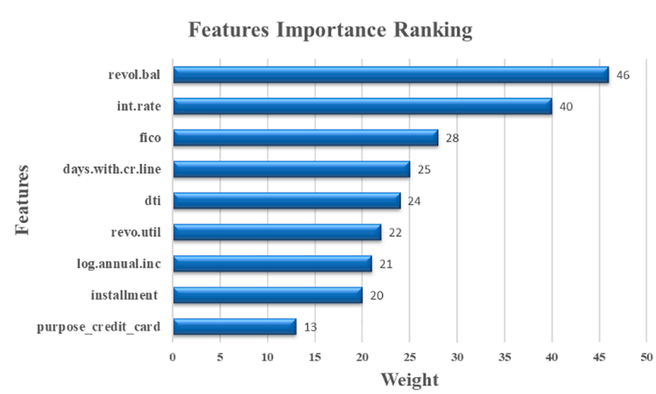

Fig. 8. Plot feature importance from "lending club loan data analysis" data.

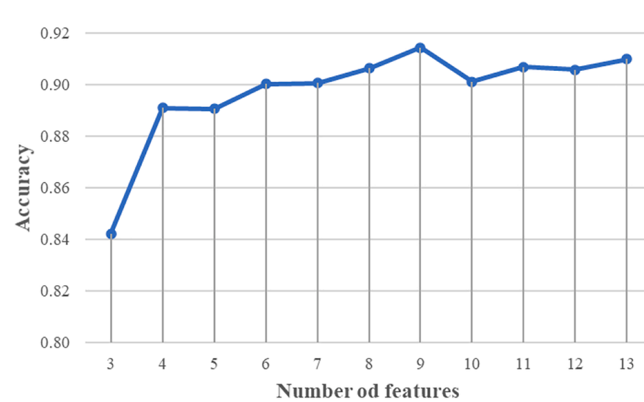

Fig. 9. Graph of feature trials on "lending club loan data analysis" data.

using the Confusion Matrix produce Accuracy, Recall and f1 scores which are shown in Tables 2 and 3.

The test results of the online P2P lending dataset in Table 2 show that the stacking ensemble model's accuracy is better than a single classifier model from KNN, SVM and Random Forest. Additionally, the model's accuracy increased after trials using different meta-learner models, namely LightGBM 99.938% AdaBoost 99.940% and the best meta-learner is XGBoost, with an accuracy of 99.945%.

After optimization with the addition of LightGBM (LGBFS) and SMOTE feature selection, the LGBFS-StackingXGBoost model was declared the best performance model with an accuracy of 99.982%, recall 0.9999 and f1 score 0.9998.

Table 3 is the test results of the lending club loan data analysis dataset. The table above shows that the stacking model's accuracy is superior to the use of the KNN, SVM and Random Forest single classifier models. Model accuracy also experienced an increase in accuracy after trials using different meta-learner models, namely AdaBoost 87.679%, LightGBM 87.857% and XGBoost is the best meta-learner with an accuracy of 87.946%.

After being optimized with LGBFS and SMOTE, the LGBFS-StackingXGBoost model was declared the best performing model with an accuracy of 91.434%, recall 0.9151 and f1 score 0.9165.

The comparatif performance of proposed method with existing similar work can be shown in Table 4.

From the results of a comparison of model performance from two different datasets, superior performance was shown in the online P2P Lending dataset trial, with superior accuracy of 99.982%. So that, our results show the stacking ensemble with XGBoost meta-learner optimized by LGBFS are the most promising in terms of the performance metrics considered. There are several knowledge contributions made to society through our research. For that, in this study we offer a computing environment on which experimental datasets could be evaluated, and gave a comparison set of ML models that may be utilized for predicting credit default risk.

5. Conclusion

In this study, a P2P lending default prediction model was created using a stacking ensemble technique that was optimized using the XGBoost, LightGBM and AdaBoost algorithms as meta-learners. The feature selection method uses the LightGBM algorithm. This study demonstrates the effectiveness of the stacking ensemble algorithm in classifying the prediction of default risk. Out of a total of 18 original features in the lending club loan data analysis dataset, only the 10 best features were selected and the online P2P lending dataset from 14 original features were selected 9 of the best features. The problem of data imbalance in the normal and default classes in the dataset is handled using the SMOTE oversampling method. The results of the evaluation test show the success of increasing the accuracy of the Stacking Ensemble technique. The best model is obtained from LGBFS-StackingXGBoost which obtains 99,982% accuracy in "online P2P lending" dataset and 91,434% accuracy in "lending club loan data analysis" dataset.

In further studies, it can be developed by conducting experiments on larger datasets or datasets from different countries and trying to tune new models to achieve better performance.

CRediT authorship contribution statement

Much Aziz Muslim: Conceptualization, Methodology, Software, Project administration. **Tiara Lailatul Nikmah:** Software, Writing – original draft. **Dwika Ananda Agustina Pertiwi:** Software, Writing – review & editing. **:** Validation. **:** Validation. **Yosza Dasril:** Supervision. **:** Validation.

Table 2
Evaluation of different methods from dataset online P2P lending.

Algorithm	Accuracy(%)	Recall	f1 Score
SVM	96,326%	0,9632	0,9813
KNN	97,841%	0,9781	0,9889
Random Forest	99,937%	0,9994	0,9996
Stacking-LightGBM	99,938%	0,9996	0,9997
Stacking-AdaBoost	99,940%	0,9995	0,9997
Stacking-XGBoost	99,945%	0,9996	0,9997
LGBFS-StackingXGBoost	**99,982%**	**0,9999**	**0,9998**

Table 3
Evaluation of different methods from dataset lending club loan data analysis.

Algorithm	Accuracy(%)	Recall	f1 Score
KNN	86,696%	0,8777	0,9286
SVM	87,768%	0,8777	0,9349
Random Forest	87,779%	0,8790	0,9347
Stacking-AdaBoost	87,679%	0,8782	0,9343
Stacking-LightGBM	87,857%	0,8791	0,9352
Stacking-XGBoost	87,946%	0,8799	0,9357
LGBFS-StackingXGBoost	**91,434%**	**0,9151**	**0,9165**

Table 4
Comparative performance of proposed method with existing similar work.

Author (year)	Data Used	Classifier	Performance
(Zhu et al., 2019)	Loan data Lending Club	Random Forest	Accuracy: 98% AUC: 98.3% Recall: 98% F1-score: 98%
(M. Li et al., 2021)	Customer Loan data Lending Club	StackingSVM+CNN	Accuracy: 96.5% AUC: 96.5% F1-score: 97.1%
(Kun et al., 2020)	LendingClub dataset-Q4 2018	StackingLR	Accuracy: 98% AUC: 98.1% Precision: 97.9% Recall: 98.6% F1-score: 98.3%
Proposed method	1. Lending Club Loan data	LGBFS+StackingXGBoost	Accuracy: 91.43% Recall: 91.51% F1-score: 91.65%
	2. OnlineP2P Lending	LGBFS+StackingXGBoost	Accuracy: 99.982% Recall: 99.99% F1-score: 99.99%

Declaration of Competing Interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

References

Abdar, M., Zomorodi-Moghadam, M., Zhou, X., Gururajan, R., Tao, X., Barua, P. D., & Gururajan, R. (2020). A new nested ensemble technique for automated diagnosis of breast cancer. *Pattern Recognition Letters, 132*, 123–131.

Al-Asadi, M. A., & Tasdemír, S. (2021). Empirical comparisons for combining balancing and feature selection strategies for characterizing football players using FIFA video game system. *IEEE Access, 9*, 149266–149286. https://doi.org/10.1109/ACCESS.2021.3124931

Anagnostou, E., Dimopoulou, P., Sklavos, S., Zouvelou, V., & Zambelis, T. (2021). Identifying jitter outliers in single fiber electromyography: Comparison of four methods. *Muscle \& Nerve, 63*(2), 217–224.

Ashari, I. A., Muslim, M. A., & Alamsyah, A. (2016). Comparison performance of genetic algorithm and ant colony optimization in course scheduling optimizing. *Scientific Journal of Informatics, 3*(2), 149–158. https://doi.org/10.15294/sji.v3i2.7911

Budiman, K., & Ifriza, Y. N. (2021). Analysis of earthquake forecasting using random forest. *Journal of Soft Computing Exploration, 2*(2), 153–162. https://www.shmpublisher.com/index.php/joscex/article/view/51.

Byanjankar, A., Mezei, J., & Heikkilä, M. (2021). Data-driven optimization of peer-to-peer lending portfolios based on the expected value framework. *Intelligent Systems in Accounting, Finance and Management, 28*(2), 119–129.

Chaudhary, N., Abu-Odeh, A., Karaman, I., & Arróyave, R. (2017). A data-driven machine learning approach to predicting stacking faulting energy in austenitic steels. *Journal of Materials Science, 52*(18), 11048–11076. https://doi.org/10.1007/s10853-017-1252-x

Chen, Y.-R., Leu, J.-S., Huang, S.-A., Wang, J.-T., & Takada, J.-I. (2021). Predicting default risk on peer-to-peer lending imbalanced datasets. *IEEE Access, 9*, 73103–73109. https://doi.org/10.1109/ACCESS.2021.3079701

Chicco, D., & Jurman, G. (2020). The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. *BMC Genomics, 21*(1), 1–13.

Damayanti, D. R., Wicaksono, S., Hakim, M. F. Al, Jumanto, J., Subhan, S., & Ifriza, Y. N (2022). Rainfall prediction in Blora regency using mamdani's fuzzy inference system. *Journal of Soft Computing Exploration, 3*(1), 62–69. https://doi.org/10.52465/joscex.v3i1.69

Devan, P., & Khare, N. (2020). An efficient XGBoost–DNN-based classification model for network intrusion detection system. *Neural Computing and Applications, 32*(16), 12499–12514.

Falasari, A., & Muslim, M. A. (2022). Optimize naïve bayes classifier using chi square and term frequency inverse document frequency for amazon review sentiment analysis. *Journal of Soft Computing Exploration, 3*(1), 31–36. https://doi.org/10.52465/joscex.v3i1.68

Gupta, A., Khan, R. U., Singh, V. K., Tanveer, M., Kumar, D., Chakraborti, A., & Pachori, R. B. (2020). A novel approach for classification of mental tasks using multiview ensemble learning (MEL). *Neurocomputing, 417*, 558–584. https://doi.org/10.1016/j.neucom.2020.07.050

Hazimah, N., Harahap, S., Amirullah, A., Saputro, M. B., & Tamaroh, I. A. (2022). Classification of potential customers using C4 . 5 and k-means algorithms to determine customer service priorities to maintain loyalty. *Journal of Soft Computing Exploration, 3*(2), 123–130. https://doi.org/10.52465/joscex.v3i2.89

Hou, W.hui, Wang, X.kang, Zhang, H.yu, Wang, J.qiang, & Li, L (2020). A novel dynamic ensemble selection classifier for an imbalanced data set: An application for credit risk assessment. *Knowledge-Based Systems, 208*, Article 106462. https://doi.org/10.1016/j.knosys.2020.106462

Jadwal, P. K., Jain, S., Pathak, S., & Agarwal, B. (2022). Improved resampling algorithm through a modified oversampling approach based on spectral clustering and SMOTE. *Microsystem Technologies, 28*(12), 2669–2677. https://doi.org/10.1007/s00542-022-05287-8

Jha, M., Gupta, R., & Saxena, R. (2022). A framework for in-vivo human brain tumor detection using image augmentation and hybrid features. *Health Information Science and Systems, 10*(1), 1–12.

Kang, S., Cho, S., & Kang, P. (2015). Multi-class classification via heterogeneous ensemble of one-class classifiers. *Engineering Applications of Artificial Intelligence, 43*, 35–43. https://doi.org/10.1016/j.engappai.2015.04.003

Khochare, J., Rathod, J., Joshi, C., & Laveti, R. N. (2020). A short-term wind forecasting framework using ensemble learning for indian weather stations. In *2020 IEEE International Conference for Innovation in Technology, INOCON 2020*. https://doi.org/10.1109/INOCON50539.2020.9298262

Kun, Z., Weibing, F., & Jianlin, W. (2020). Default identification of P2P lending based on stacking ensemble learning. In *2020 2nd International Conference on Economic Management and Model Engineering (ICEMME)* (pp. 992–1006). https://doi.org/10.1109/ICEMME51517.2020.00203

Lee, D., & Kim, K. (2021). An efficient method to determine sample size in oversampling based on classification complexity for imbalanced data. *Expert Systems with Applications, 184*, Article 115442. https://doi.org/10.1016/j.eswa.2021.115442

Li, L.-H., Sharma, A. K., Ahmad, R., & Chen, R.-C. (2021). Predicting the default borrowers in P2P platform using machine learning models. In A. Solanki, S. K. Sharma, S. Tarar, P. Tomar, S. Sharma, & A. Nayyar (Eds.), *Artificial Intelligence and Sustainable Computing for Smart City* (pp. 267–281). Springer International Publishing.

Li, M., Yan, C., & Liu, W. (2021). The network loan risk prediction model based on Convolutional neural network and Stacking fusion model. *Applied Soft Computing, 113*, Article 107961. https://doi.org/10.1016/j.asoc.2021.107961

Li, W., Ding, S., Wang, H., Chen, Y., & Yang, S. (2020). Heterogeneous ensemble learning with feature engineering for default prediction in peer-to-peer lending in China. *World Wide Web, 23*(1), 23–45. https://doi.org/10.1007/s11280-019-00676-y

Li, Z.-S., Yao, X., Liu, Z.-G., & Zhang, J.-C. (2021). Feature Selection Algorithm Based on LightGBM; [基于LightGBM的特征选择算法]. *Dongbei Daxue Xuebao/Journal of Northeastern University, 42*(12), 1688–1695. https://doi.org/10.12068/j.issn.1005-3026.2021.12.003

Liang, K., Zhang, C., & Jiang, C. (2022). Analyzing default risk among P2P platforms based on the LAS-STACK method by considering multidimensional signals under specific economic contexts. *Electronic Commerce Research, 22*(1), 77–111. https://doi.org/10.1007/s10660-021-09505-9

Liang, L., & Cai, X. (2020). Forecasting peer-to-peer platform default rate with LSTM neural network. *Electronic Commerce Research and Applications, 43*, Article 100997. https://doi.org/10.1016/j.elerap.2020.100997

Lopez-Arevalo, I., Aldana-Bobadilla, E., Molina-Villegas, A., Galeana-Zapién, H., Muñiz-Sanchez, V., & Gausin-Valle, S. (2020). A memory-efficient encoding method for processing mixed-type data on machine learning. *Entropy, 22*(12), 1391.

Luo, H., & Yan, D. (2022). Blockchain architecture and its applications in a bank risk mitigation framework. *Economic Research-Ekonomska Istraživanja, 35*(1), 3119–3137.

Ma, X., Sha, J., Wang, D., Yu, Y., Yang, Q., & Niu, X. (2018). 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. *Electronic Commerce Research and Applications, 31*, 24–39. https://doi.org/10.1016/j.elerap.2018.08.002

Machado, M. R., Karray, S., & de Sousa, I. T. (2019). LightGBM: An effective decision tree gradient boosting method to predict customer loyalty in the finance industry. In *2019 14th International Conference on Computer Science & Education (ICCSE)* (pp. 1111–1116). https://doi.org/10.1109/ICCSE.2019.8845529

Mardiansyah, M. F., Pratama, R., Hakim, M. F. Al, & Rawat, B (2022). Optimization of breast cancer classification using feature selection on neural network. *Journal of Soft Computing Exploration, 3*(2), 105–110. https://doi.org/10.52465/joscex.v3i2.78

Mienye, I. D., & Sun, Y. (2021). Performance analysis of cost-sensitive learning methods with application to imbalanced medical data. *Informatics in Medicine Unlocked, 25*, Article 100690. https://doi.org/10.1016/j.imu.2021.100690

Mohammadi, S., Mirvaziri, H., Ghazizadeh-Ahsaee, M., & Karimipour, H. (2019). Cyber intrusion detection by combined feature selection algorithm. *Journal of Information Security and Applications, 44*, 80–88.

Mohammed, A., & Kora, R. (2022). An effective ensemble deep learning framework for text classification. *Journal of King Saud University - Computer and Information Sciences, 34*(10), 8825–8837. https://doi.org/10.1016/j.jksuci.2021.11.001. Part A.

Mota, J., Moreira, A. C., & Brandão, C. (2018). Determinants of microcredit repayment in Portugal: Analysis of borrowers, loans and business projects. *Portuguese Economic Journal, 17*(3), 141–171.

Mukherjee, M., & Khushi, M. (2021). SMOTE-ENC: A novel SMOTE-based method to generate synthetic data for nominal and continuous features. *Applied System Innovation, 4*(1), 18.

Muslim, M A, Dasril, Y., Alamsyah, A., & Mustaqim, T. (2021). Bank predictions for prospective long-term deposit investors using machine learning LightGBM and SMOTE. *Journal of Physics: Conference Series, 1918*(4). https://doi.org/10.1088/1742-6596/1918/4/042143

Muslim, M A, Herowati, A. J., Sugiharti, E., & Prasetiyo, B. (2018). Application of the pessimistic pruning to increase the accuracy of C4.5 algorithm in diagnosing chronic kidney disease. *Journal of Physics: Conference Series, 983*(1). https://doi.org/10.1088/1742-6596/983/1/012062

Muslim, Much Aziz, Dasril, Y., Sam'an, M., & Ifriza, Y. N (2022). An improved light gradient boosting machine algorithm based on swarm algorithms for predicting loan default of peer-to-peer lending. *Indonesian Journal of Electrical Engineering and Computer Science, 28*(2), 1002–1011. https://doi.org/10.11591/ijeecs.v28.i2.pp1002-1011

Mustaqim, T., Umam, K., & Muslim, M. A. (2020). Twitter text mining for sentiment analysis on government's response to forest fires with vader lexicon polarity detection and k-nearest neighbor algorithm. *Journal of Physics: Conference Series*, (3), 1567. https://doi.org/10.1088/1742-6596/1567/3/032024

Papouskova, M., & Hajek, P. (2019). Two-stage consumer credit risk modelling using heterogeneous ensemble learning. *Decision Support Systems, 118*, 33–45. https://doi.org/10.1016/j.dss.2019.01.002

Perez, C., Sokolova, K., & Konate, M. (2020). Digital social capital and performance of initial coin offerings. *Technological Forecasting and Social Change, 152*, Article 119888. https://doi.org/10.1016/j.techfore.2019.119888

Pradipta, G. A., Wardoyo, R., Musdholifah, A., & Sanjaya, I. N. H. (2021). Radius-SMOTE: A new oversampling technique of minority samples based on radius distance for learning from imbalanced data. *IEEE Access, 9*, 74763–74777. https://doi.org/10.1109/ACCESS.2021.3080316

Prasetiyo, B., Alamsyah Muslim, M. A., & Baroroh, N (2021a). Evaluation of feature selection using information gain and gain ratio on bank marketing classification using naïve bayes. *Journal of Physics: Conference Series, 1918*(4). https://doi.org/10.1088/1742-6596/1918/4/042153

Prasetiyo, B., Alamsyah Muslim, M. A., & Baroroh, N (2021b). Evaluation performance recall and F2 score of credit card fraud detection unbalanced dataset using SMOTE oversampling technique. *Journal of Physics: Conference Series, 1918*(4). https://doi.org/10.1088/1742-6596/1918/4/042002

Prasetiyo, B., Alamsyah, Muslim, Subhan, M. A., & Baroroh, N (2020). Artificial neural network model for banckrupty prediction. *Journal of Physics: Conference Series*, (3), 1567. https://doi.org/10.1088/1742-6596/1567/3/032022

Ragab, M., Abdel Aal, A. M. K., Jifri, A. O., & Omran, N. F (2021). Enhancement of predicting students performance model using ensemble approaches and educational data mining techniques. *Wireless Communications and Mobile Computing*, 2021.

Ruyu, B., Mo, H., & Haifeng, L. (2019). A Comparison of Credit Rating Classification Models Based on Spark- Evidence from Lending-club. *Procedia Computer Science, 162*, 811–818. https://doi.org/10.1016/j.procs.2019.12.054 (Itqm).

Sagar, A., Vega, C., Bouriaud, O., Piedallu, C., & Renaud, J.-P. (2022). Multisource forest inventories: A model-based approach using k-NN to reconcile forest attributes statistics and map products. *ISPRS Journal of Photogrammetry and Remote Sensing, 192*, 175–188. https://doi.org/10.1016/j.isprsjprs.2022.08.016

Sarıkaya, A., Günel Kılıç, B., & Demirci, M. (2022). GRU-GBM: A combined intrusion detection model using LightGBM and gated recurrent unit. *Expert Systems, 39*(9), e13067. https://doi.org/10.1111/exsy.13067

Shorfuzzaman, M. (2021). IoT-enabled stacked ensemble of deep neural networks for the diagnosis of COVID-19 using chest CT scans. *Computing*, 1–22.

Song, J., Wang, Y., Fang, Z., Peng, L., & Hong, H. (2020). Potential of ensemble learning to improve tree-based classifiers for landslide susceptibility mapping. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 13*, 4642–4662. https://doi.org/10.1109/JSTARS.2020.3014143

Song, Y., Wang, Y., Ye, X., Wang, D., Yin, Y., & Wang, Y. (2020). Multi-view ensemble learning based on distance-to-model and adaptive clustering for imbalanced credit risk assessment in P2P lending. *Information Sciences, 525*, 182–204. https://doi.org/10.1016/j.ins.2020.03.027

Susan, S., Kumar, A., & Jain, A. (2021). Evaluating heterogeneous ensembles with boosting meta-learner. *Inventive Communication and Computational Technologies* (pp. 699–710). Springer.

Vianita, E., Wibowo, A., Surarso, B., & Widodo, A. P. (2021). Car insurance segmentation prediction based on the most influential features using random forest and stacking ensemble learning. *Journal of Soft Computing Exploration, 2*(2), 86–92. https://doi.org/10.52465/joscex.v2i2.39

Wang, K., Li, M., Cheng, J., Zhou, X., & Li, G. (2022). Research on personal credit risk evaluation based on XGBoost. *Procedia Computer Science, 199*, 1128–1135. https://doi.org/10.1016/j.procs.2022.01.143

Wang, T., & Li, J. (2019). An improved support vector machine and its application in P2P lending personal credit scoring. *IOP Conference Series: Materials Science and Engineering, 490*(6), 62041. https://doi.org/10.1088/1757-899X/490/6/062041

Wang, Z., Jiang, C., Ding, Y., Lv, X., & Liu, Y. (2017). A novel behavioral scoring model for estimating probability of default over time in Peer-to-Peer lending. *Electronic Commerce Research and Applications*. https://doi.org/10.1016/j.elerap.2017.12.006

Wardoyo, R., Wirawan, I. M. A., & Pradipta, I. G. A. (2022). Oversampling Approach Using Radius-SMOTE for Imbalance Electroencephalography Datasets. *Emerging Science Journal, 6*(2), 382–398.

Xia, Y., Zhao, J., He, L., Li, Y., & Yang, X. (2021). Forecasting loss given default for peer-to-peer loans via heterogeneous stacking ensemble approach. *International Journal of Forecasting, 37*(4), 1590–1613. https://doi.org/10.1016/j.ijforecast.2021.03.002

Xiong, Yi, Wang, Q., Yang, J., Zhu, X., & Wei, D.-Q. (2018). PredT4SE-stack: prediction of bacterial type IV secreted effectors from protein sequences using a stacked ensemble method. *Frontiers in Microbiology, 9*, 2571.

Xiong, Yueling, Ye, M., & Wu, C. (2021). Cancer classification with a cost-sensitive naive bayes stacking ensemble. *Computational and Mathematical Methods in Medicine, 2021*, Article 5556992. https://doi.org/10.1155/2021/5556992

Yang, H., Luo, Y., Ren, X., Wu, M., He, X., Peng, B., Deng, K., Yan, D., Tang, H., & Lin, H. (2021). Risk prediction of diabetes: Big data mining with fusion of multifarious physical examination indicators. *Information Fusion, 75*, 140–149. https://doi.org/10.1016/j.inffus.2021.02.015

Yoon, Y., Li, Y., & Feng, Y. (2019). Factors affecting platform default risk in online peer-to-peer (P2P) lending business: An empirical study using Chinese online P2P platform data. *Electronic Commerce Research, 19*(1), 131–158. https://doi.org/10.1007/s10660-018-9291-1

Zhang, Y., Jiang, Z., Chen, C., Wei, Q., Gu, H., & Yu, B. (2022). DeepStack-DTIs: Predicting drug–target interactions using LightGBM feature selection and deep-stacked ensemble classifier. *Interdisciplinary Sciences: Computational Life Sciences, 14*(2), 311–330. https://doi.org/10.1007/s12539-021-00488-7

Zhao, H., Liu, Q., Zhu, H., Ge, Y., Chen, E., Zhu, Y., & Du, J. (2018). A sequential approach to market state modeling and analysis in online P2P lending. *IEEE Transactions on Systems, Man, and Cybernetics: Systems, 48*(1), 21–33. https://doi.org/10.1109/TSMC.2017.2665038

Zhou, J., Li, W., Wang, J., Ding, S., & Xia, C. (2019). Default prediction in P2P lending from high-dimensional data based on machine learning. *Physica A: Statistical Mechanics and Its Applications, 534*, Article 122370. https://doi.org/10.1016/j.physa.2019.122370

Zhu, L., Qiu, D., Ergu, D., Ying, C., & Liu, K. (2019). A study on predicting loan default based on the random forest algorithm. *Procedia Computer Science, 162*(Itqm 2019), 503–513. https://doi.org/10.1016/j.procs.2019.12.017