Bulletin of Electrical Engineering and Informatics
Vol. 13, No. 5, October 2024, pp. 3609∼3615
ISSN: 2302-9285, DOI: 10.11591/eei.v13i5.7651

Feature selection in P2P lending for default prediction using grey wolf optimization and machine learning

Muhammad Sam'an, Safuan, Muhamma Munsarif
Department of Informatics, Faculty of Engineering and Computer Science, Universitas Muhammadiyah Semarang, Semarang, Indonesia

Article Info

Article history:

Received Oct 1, 2023
Revised Mar 6, 2024
Accepted Mar 31, 2024

Keywords:

Evaluations
Feature selection
Grey wolf optimization
Machine learning models
Peer-to-peer lending

ABSTRACT

Online loan services like peer-to-peer (P2P) lending enable lenders to transact without bank intermediaries. Predicting which lenders are likely to default is crucial to avoid bankruptcy since lenders bear the risk of default. However, this task becomes challenging when the P2P lending dataset contains numerous features. The prediction performance could be improved if the dataset features are relevant. Hence, applying feature selection to remove redundant and irrelevant features is essential. This paper introduces a novel wrapper feature selection model to identify the optimal feature subset for predicting defaults in P2P lending. The proposed method includes two main phases: feature selection and classification. Initially, grey wolf optimization (GWO) is used to select the best features in P2P lending datasets. Then, the fitness function of GWO is assessed using ten different machine learning (ML) models. Experimental results indicate that the proposed model outperforms previous related work, achieving average accuracy, recall, precision, and F1-score of 96.77%, 80.73%, 97.52%, and 80.06%, respectively.

Corresponding Author:

Muhammad Sam'an
Department of Informatics, Faculty of Engineering and Computer Science
Universitas Muhammadiyah Semarang
Semarang, Indonesia
Email: muhammad92sam@unimus.ac.id

1. INTRODUCTION

Peer-to-peer (P2P) lending, the world's largest online credit market, offers independent loans and may influence traditional banks, prompting researchers to investigate various evaluation models. Different approaches have been suggested for credit risk assessment in P2P lending, such as stacking ensemble models, logistic and linear regression, statistical models (logistic regression (LR)), Bayesian classifiers, latent Dirichlet allocation (LDA)), and artificial intelligence models (decision tree (DT), random forest (RF), light gradient boosting machine (LightGBM), artificial neural networks (ANN), and convolutional neural network (CNN)) [1]–[5]. Additional methods include multi-view neural networks [6], [7], benchmarking models [8], ensemble resampling [7], [9], Naive Bayes and DT [10], and an enhanced DT [11] model for default prediction. Credit risk assessment has also been executed using a regression model [12], a sequential ensemble learning model [13], an ANN-based model for credit scoring [14], and an RF model for borrower identification [15]. Outlier detection relies on poor credit scores [16], while internal rate of return is utilized to forecast expected profits [17].

P2P lending datasets often contain inappropriate and redundant features, impacting predictive model accuracy [18]. The large dataset dimensions hinder efficient model performance, requiring prolonged processing times. Feature selection improves accuracy by identifying relevant features and reducing data dimensions [19]–[21]. Various methods, including max-relevance and min-redundancy (MRMR), coupled with k-means clustering, LightGBM, RF, LR, RF, and extreme gradient boosting (XGBoost), binary particle swarm optimization (PSO) with support vector machine (SVM) , adaptive feature selection, most informative graph, most relative graph, and grey relational clustering, have been proposed for predicting P2P loan defaults. These models consistently demonstrate superior accuracy compared to those without feature selection [1], [2], [22]–[28].

The traditional feature selection technique needs to be updated to address the issue, as it significantly expands the feature space, making it computationally difficult [29]. Evaluating 2^N subsets is known to be an NP-hard problem [21], [22], leading to the adoption of evolutionary algorithms (EA) as an alternative solution. Grey wolf optimization (GWO), introduced by [30], is an effective EA for optimizing feature selection in large feature spaces. This paper proposes a feature selection method based on GWO and machine learning (ML) models to achieve highly accurate loan default predictions using data from P2P lending platforms.

2. METHOD

Figure 1 visualizes the workflow of research method, namely feature selection based on GWO and 10-ML for predicting default in P2P lending.

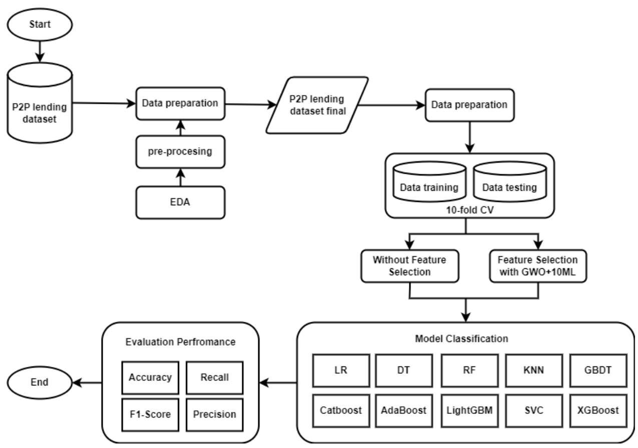

Figure 1. The process of the research method

2.1. Pre-processing

The dataset used in this study was obtained from the Kaggle repository, specifically the online P2P lending dataset [31]. It comprises Lending Club loan data from 2023, featuring 16,000 user loans and 28 attributes, including loan_amnt, term, emp_length, annual_'inc, dti, delinq_2yrs, fico_range_low, inq_last_6mths, open_acc, pub_rec, revol_bal, revol_util, total_acc, installmen, int_rate, out_prncp, last_pymnt_amnt, total_rec_int, total_rec_late_fee, recoveries, collection_recover_fee, and default.

No missing values were identified in the P2P lending dataset. We excluded features associated with outliers, such as 'grade' and 'sub grade,' and performed one-hot encoding on categorical variables like 'term,' 'home ownership,' 'purpose,' 'verification status,' and 'pymnt plan.' The final dataset for analysis includes 41 independent features, with 'default' as the target variable.

2.2. Data splitting: stratified k-fold validation

A stratified k-fold validation with k=10 was applied to the P2P lending dataset, specifically focusing on "fully paid" and "charged off" loan status categories. The dataset was divided into 10 folds, ensuring equal representation of both categories in each fold. This approach facilitated a comprehensive evaluation of the ML model's performance across diverse data subsets, reducing potential bias and ensuring reliable assessment of its predictive capabilities.

2.3. Proposed model

This paper presents an efficient feature selection method that combines GWO with 10 machine learning models (RF, DT, support vector classifier (SVC), k-nearest neighbor (KNN), gradient boosting decision tree (GBDT), LightGBM, XGBoost, adaptive boosting (AdaBoost), categorical boosting (CatBoost), and LR). The method operates in two main phases: first, GWO removes redundant and irrelevant features, then binary classification is performed using the optimized feature set. GWO efficiently processes the P2P lending dataset, resulting in high classification accuracy with a reduced number of features. Inspired by the social hierarchy and group hunting behavior of grey wolves, our approach mathematically models the primary phases of the hunt—tracking, chasing, and attacking prey—for the GWO design and optimizations, represented by α, β, δ, and ϖ. In (1) and (2) simulate the circular behavior of surrounding prey.

$$\vec{Z}(t+1) = \vec{Z}_k(t) + \vec{P} \cdot \vec{R} \tag{1}$$

$$\vec{R} = |\vec{Q} \cdot \vec{Z}_k(t) - \vec{Z}(t)| \tag{2}$$

where $\vec{D}$ is defined in (2), t is the number of iterations. $\vec{P}$ and $\vec{R}$ are the coefficient vectors, $\vec{Z}_k$ is the position of the prey, $\vec{Z}$ is the position of the grey wolf. $\vec{P}$ and $\vec{R}$ are determined using (3) and (4):

$$\vec{P} = 2p \cdot \vec{s}_1 - p \tag{3}$$

$$\vec{R} = 2\vec{s}_2 \tag{4}$$

where 's' is linearly decreasing from 2 to 0 during the iteration. $\vec{r}_1, \vec{r}_2$ are random vector $[0, 1]$ so that the grey wolf's position is updated as in (5):

$$\vec{Z}(t+1) = \frac{\vec{Z}_1 + \vec{Z}_2 + \vec{Z}_3}{3} \tag{5}$$

$$\begin{array}{rcl} \vec{Z}_1 & = & \vec{Z}_\alpha - \vec{Z}_1 \cdot \vec{R}_\alpha \\ \vec{Z}_2 & = & \vec{Z}_\beta - \vec{Z}_2 \cdot \vec{R}_\beta \\ \vec{Z}_3 & = & \vec{Z}_\delta - \vec{Z}_3 \cdot \vec{R}_\delta \end{array} \tag{6}$$

where $\vec{Z}_\alpha, \vec{Z}_\beta, \vec{Z}_\delta$ are three initial best solutions in the swarm at a given iteration. $\vec{P}_1, \vec{P}_2, \vec{P}_3$ are defined in (3): $\vec{R}_\alpha, \vec{R}_\beta, \vec{R}_\delta$ are defined in (7):

$$\begin{array}{rcl} \vec{R}_\alpha & = & |\vec{Q}_1 \cdot \vec{Z}_\alpha - \vec{Z}| \\ \vec{R}_\beta & = & |\vec{Q}_2 \cdot \vec{Z}_\beta - \vec{Z}| \\ \vec{R}_\delta & = & |\vec{Q}_3 \cdot \vec{Z}_\delta - \vec{Z}| \end{array} \tag{7}$$

where $\vec{Q}_1, \vec{Q}_2, \vec{Q}_3$ are defined in (4). Thus, GWO is an update of the 'p' parameter, which controls the trade-off between exploration and exploitation. Parameter 'p' is updated linearly in each iteration with a range of 2 to 0 according to (8):

$$p = 2 - t(\frac{2}{\max iter}) \tag{8}$$

$\max iter$ is the total number of iterations allowed for optimization:

$$fitness(t) = \alpha\rho + \beta(\frac{N-L}{L}) \tag{9}$$

ρ denotes the classification accuracy of the 10-ML models. L refers to the number of selected features, while N represents the total number of features in the P2P lending dataset. The parameters α and β correspond to the weights assigned to classification accuracy and feature selection quality, respectively, with $\alpha \in [0, 1]$ and $\beta = 1 - \alpha$. The variable t signifies the iteration number.

3. RESULT AND DISCUSSION

The list of the most influential features is generated through the feature selection process using GWO with 10-ML models on the P2P lending dataset as input. The rationale behind choosing 95 wolves as the population size in GWO is to enhance the diversity of the solution search, enabling broader coverage in the search space. The substantial number of wolves is expected to increase the probability of finding an optimal subset of features. Other parameter settings include 10 iterations, 41 features or dimensions, and a search domain of [0 1]. Additionally, the alpha and beta values for the fitness function are set at 0.4999 and 0.5, respectively. This approach ensures that the feature selection process is performed optimally, yielding the best subset of features in P2P lending. The proposed of GWO+10-ML model follows the workflow outlined in Algorithm 1, and its output is represented by the fitness function displayed in Figure 2.

Algorithm 1 . Pseudo-code of GWO

Input: the prey wolf population Z_i (1,2, . . ., n)
Output: $\vec{Z}_\alpha$ is the best search agent, $\vec{Z}_\beta$ is the second best search agent, $\vec{Z}_\delta$ is third best agent
Initialization: a, $\vec{P}$ and $\vec{Q}$
Calculate the fitness of each search agent by Eq. (9)
while true **do**
    t $< \max iter$
    for all search agent **do**
        update the position by Eq. (5)
    update a, $\vec{P}$ and $\vec{Q}$
    update $\vec{Z}_\alpha, \vec{Z}_\beta, \vec{Z}_\delta$
    t=t+1
end
return $\vec{Z}_\alpha$

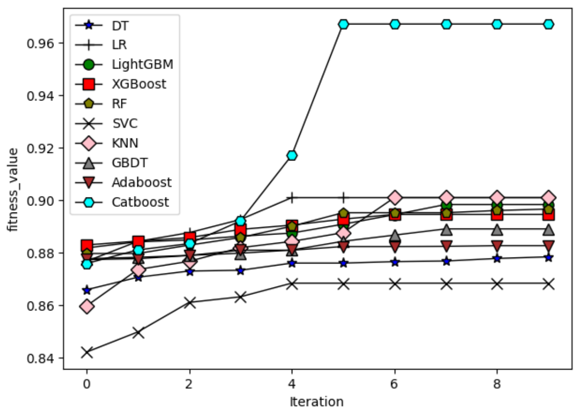

Figure 2. The fitness value of the proposed model

The fitness value consistently increases with each iteration, indicating that GWO effectively identifies optimal features for the 10-ML classification process. Figure 3 visually represents the selected feature-length (L), denoting the optimal features chosen by the 10-ML models. Additionally, the computational time for achieving the optimal solution in GWO+10-ML models varies. Figure 3 provides insights into the time consumption during iterations for calculating fitness values, revealing that GWO+CatBoost requires less time than other models. It is important to note that faster iteration speed does not necessarily guarantee superior prediction accuracy compared to other models.

Reducing the number of features in the dataset is crucial for addressing the "curse of dimensionality", improving model learning, and avoiding overfitting. This process streamlines model training by eliminating irrelevant or redundant features, enhancing performance, and simplifying interpretation. Performance evaluation, detailed in Table 1, indicates significant differences between GWO+10-ML and original 10-ML models. GWO+10-ML consistently outperforms across accuracy, recall, precision, and F1-score, with averages of 96.77%, 80.73%, 97.52%, and 80.06%, respectively, compared to lower averages from original 10-ML models. These results affirm the superiority of GWO+10-ML models and emphasize the effectiveness of feature optimization, particularly in GWO+KNN and GWO+LightGBM models, in improving predictive performance. This underscores the importance of efficient feature selection in identifying potential defaulters in the P2P lending environment.

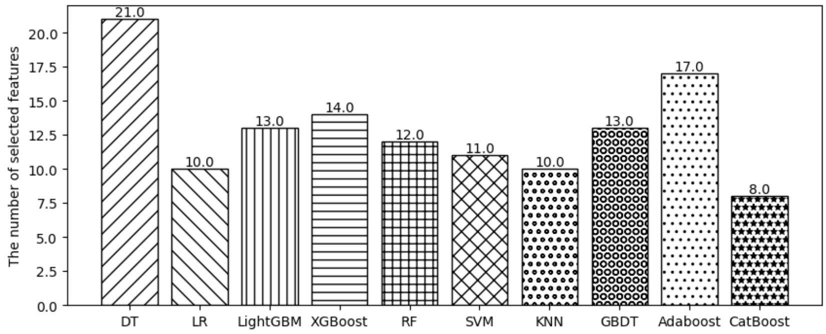

Figure 3. The length of the selected features (L)

Table 1. The evaluation performance comparison of GWO+10-ML and original ML

Models	Accuracy (%)		Recall (%)		Precision (%)		F1-score (%)	
	Original	GWO	Original	GWO	Original	GWO	Original	GWO
DT	95.78	96.63	84.97	87.89	86.60	89.38	85.77	88.63
LR	94.84	96.41	70.77	75.99	93.13	100.00	80.43	86.36
LightGBM	97.44	97.97	85.80	88.94	96.71	97.26	90.93	92.91
XGBoost	97.50	97.81	86.22	88.31	96,72	96.80	91.17	92.36
RF	96.53	97.31	76.83	85.39	100	96.24	86.89	90.49
SVM	92.31	94.94	52.19	66.18	93.63	100	67.02	79.65
KNN	86.72	96.41	17.95	75.99	72.88	100	28.81	86.36
GBDT	96.75	96.94	79.96	81.63	97.95	97.51	88.05	88.86
AdaBoost	96.88	96.97	81.84	81.42	96.79	97.99	88.69	88.94
CatBoost	96,34	96.34	75.57	75.57	100.00	100.00	86.09	86.09
Average	95.11	96.77	71.21	80.73	93.44	97.52	79.39	88.06

Moreover, Table 2 (see in Appendix) provides a comparative assessment of the accuracy performance of the proposed model in relation to prior related research. The table demonstrates the superior performance of the proposed model compared to the studies conducted by Truong *et al.* [32], Setiawan *et al.* [25], as well as Victor and Raheem [33]. Consequently, it reaffirms the effectiveness of GWO as an optimization algorithm for feature selection. Nevertheless, further development of GWO is warranted, particularly by expanding its search capabilities to accommodate high-dimensional dataset.

4. CONCLUSION

The GWO+10-ML models (including RF, DT, SVM, KNN, GBDT, LightGBM, XGBoost, AdaBoost, CatBoost, and LR) presented in this paper act as a wrapper feature selection method. These models efficiently select relevant features and discard irrelevant ones in the P2P lending dataset. Performance evaluations (accuracy, recall, precision, and F1-score) show that the GWO+10-ML models outperform the original 10-ML models in predicting defaults in P2P lending. Additionally, this method surpasses the accuracy of three previous studies. Enhancing GWO by expanding the search space is essential for handling high-dimensional datasets.

APPENDIX

Table 2. The comparison of the proposed model with the previous related work

Study	Feature selection	Models	Accuracy (%)
[32]	Restricted Boltzmann Machine	LDA	81.20
		LR	81.05
		ANN	66.05
		KNN	72.55
		SVM	76.56
		RF	67.72
[25]	Binary PSO	Electrical resistivity tomography (ERT)	64
[33]	Genetic algorithm	LR	86
		RF	92
		SVM	85
Proposed model	GWO	DT	96.63
		LR	96.41
		LightGBM	97.97
		XGBoost	97.81
		RF	97.31
		SVM	94.94
		KNN	96.41
		GBDT	96.94
		AdaBoost	96.97
		CatBoost	96.34

REFERENCES

[1] W. Yin, B. Kirkulak-Uludag, D. Zhu, and Z. Zhou, "Stacking ensemble method for personal credit risk assessment in Peer-to-Peer lending," *Applied Soft Computing*, vol. 142, 2023, doi: 10.1016/j.asoc.2023.110302.

[2] 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, 2023, doi: 10.1016/j.iswa.2023.200204.

[3] M. Munsarif, M. Sam'an, and Safuan, "Peer to peer lending risk analysis based on embedded technique and stacking ensemble learning," *Bulletin of Electrical Engineering and Informatics*, vol. 11, no. 6, pp. 3483–3489, 2022, doi: 10.11591/eei.v11i6.3927.

[4] Y. Rong, S. Liu, S. Yan, W. W. Huang, and Y. Chen, "Proposing a new loan recommendation framework for loan allocation strategies in online P2P lending," *Industrial Management and Data Systems*, vol. 123, no. 3, pp. 910–930, 2023, doi: 10.1108/IMDS-07-2022-0399.

[5] P. C. Ko, P. C. Lin, H. T. Do, and Y. F. Huang, "P2P Lending Default Prediction Based on AI and Statistical Models," *Entropy*, vol. 24, no. 6, 2022, doi: 10.3390/e24060801.

[6] Y. Tan and G. Zhao, "Multi-view representation learning with Kolmogorov-Smirnov to predict default based on imbalanced and complex dataset," *Information Sciences*, vol. 596, pp. 380–394, 2022, doi: 10.1016/j.ins.2022.03.022.

[7] Y. Song, Y. Wang, X. Ye, D. Wang, Y. Yin, and Y. Wang, "Multi-view ensemble learning based on distance-to-model and adaptive clustering for imbalanced credit risk assessment in P2P lending," *Inormation Sciences*, vol. 525, pp. 182–204, 2020, doi: 10.1016/j.ins.2020.03.027.

[8] V. Moscato, A. Picariello, and G. Sperlí, "A benchmark of machine learning approaches for credit score prediction," *Expert Systems with Applications*, vol. 165, 2021, doi: 10.1016/j.eswa.2020.113986.

[9] K. Niu, Z. Zhang, Y. Liu, and R. Li, "Resampling ensemble model based on data distribution for imbalanced credit risk evaluation in P2P lending," *Information Sciences*, vol. 536, pp. 120–134, 2020, doi: 10.1016/j.ins.2020.05.040.

[10] M. J. Christ, R. N. P. Tri, W. Chandra, and T. Mauritsius, "Lending club default prediction using Naïve Bayes and decision tree," *International Journal of Advanced Trends in Computer Science and Engineering*, vol. 8, no. 5, pp. 2528–2534, 2019, doi: 10.30534/ijatese/2019/99852019.

[11] A. Semiu and A. A. R. Gilal, "A boosted decision tree model for predicting loan default in P2P lending communities," *International Journal of Engineering and Advanced Technology*, vol. 9, no. 1, pp. 1257–1261, 2019, doi: 10.35940/ijeat.A9626.109119.

[12] S. F. Chen, G. Charkaborty, L. H. Li, and C. T. Lin, "Credit risk assessment using regression model on P2P lending," *International Journal of Applied Science and Engineering*, vol. 16, no. 2, pp. 149–157, 2019, doi: 10.6703/IJASE.201909.

[13] W. Li, S. Ding, H. Wang, Y. Chen, and S. Yang, "Heterogeneous ensemble learning with feature engineering for default prediction in peer-to-peer lending in China," *World Wide Web*, vol. 23, no. 1, pp. 23–45, 2020, doi: 10.1007/s11280-019-00676-y.

[14] A. Byanjankar, M. Heikkila, and J. Mezei, "Predicting credit risk in peer-to-peer lending: A neural network approach," in *Proceedings-2015 IEEE Symposium Series on Computational Intelligence, SSCI 2015*, 2015, pp. 719–725, doi: 10.1109/SSCI.2015.109.

[15] M. Malekipirbazari and V. Aksakalli, "Risk assessment in social lending via random forests," *Expert Systems with Applications.*, vol. 42, no. 10, pp. 4621–4631, 2015, doi: 10.1016/j.eswa.2015.02.001.

[16] H. Li, Y. Zhang, N. Zhang, and H. Jia, "Detecting the Abnormal Lenders from P2P Lending Data," in *Procedia Computer Science*, 2016, vol. 91, pp. 357–361. doi: 10.1016/j.procs.2016.07.095.

[17] C. Serrano-Cinca and B. Gutiérrez-Nieto, "The use of profit scoring as an alternative to credit scoring systems in peer-to-peer (P2P) lending," *Decision support systems*, vol. 89, pp. 113–122, 2016, doi: 10.1016/j.dss.2016.06.014.

[18] R. C. Chen, C. Dewi, S. W. Huang, and R. E. Caraka, "Selecting critical features for data classification based on machine learning methods," *Journal of Big Data*, vol. 7, no. 1, 2020, doi: 10.1186/s40537-020-00327-4.

[19] H. D. Wang, "Research on the features of car insurance data based on machine learning," in *Procedia Computer Science*, 2020, vol. 166, pp. 582–587. doi: 10.1016/j.procs.2020.02.016.
[20] M. Papouskova and P. Hajek, "Two-stage consumer credit risk modelling using heterogeneous ensemble learning," *Decision support systems*, vol. 118, pp. 33–45, 2019, doi: 10.1016/j.dss.2019.01.002.
[21] S. Gu, R. Cheng, and Y. Jin, "Feature selection for high-dimensional classification using a competitive swarm optimizer," *Soft Computing*, vol. 22, no. 3, pp. 811–822, 2018, doi: 10.1007/s00500-016-2385-6.
[22] A. K. Sharma, L. H. Li, and R. Ahmad, "Default Risk Prediction Using Random Forest and XGBoosting Classifier," in *Smart Innovation, Systems, and Technologies*, 2023, vol. 314, pp. 91–101. doi: 10.1007/978-3-031-05491-4_10.
[23] J. Xu, D. Chen, and M. Chau, "Identifying features for detecting fraudulent loan requests on P2P platforms," in *IEEE International Conference on Intelligence and Security Informatics: Cybersecurity and Big Data, ISI 2016*, 2016, pp. 79–84, doi: 10.1109/ISI.2016.7745447.
[24] S. Chen, G. Chakraborty, and L. Li, "Feature Selection on Credit Risk Prediction for Peer-to-Peer Lending," in *International Symposium on Artificial Intelligence*, 2019, pp. 5–18.
[25] N. Setiawan, Suharjito, and Diana, "A comparison of prediction methods for credit default on peer to peer lending using machine learning," in *Procedia Computer Science*, 2019, vol. 157, pp. 38–45. doi: 10.1016/j.procs.2019.08.139.
[26] L. Cui, Y. Jiao, L. Bai, L. Rossi, and E. R. Hancock, "Adaptive feature selection based on the most informative graph-based features," in *Lecture Notes in Computer Science*, 2017, vol. 10310, pp. 276–287, doi: 10.1007/978-3-319-58961-9_25.
[27] L. Cui, L. Bai, Y. Wang, X. Bai, Z. Zhang, and E. R. Hancock, "P2P lending analysis using the most relevant graph-based features," in *Lecture Notes in Computer Science*, vol. 10029, pp. 3–14, 2016, doi: 10.1007/978-3-319-49055-7_1.
[28] L. He, H. Xu, and G. Y. Ke, "A hybrid predictive framework for evaluating P2P credit risks," *Grey Systems: Theory and Application*, vol. 12, no. 3, pp. 551–573, 2022, doi: 10.1108/GS-03-2021-0041.
[29] C. Shen and K. Zhang, "Two-stage improved Grey Wolf optimization algorithm for feature selection on high-dimensional classification," *Complex and Intelligent Systems*, vol. 8, no. 4, pp. 2769–2789, 2022, doi: 10.1007/s40747-021-00452-4.
[30] S. Mirjalili, S. M. Mirjalili, and A. Lewis, "Grey Wolf Optimizer," *Advances in Engineering Software*, vol. 69, pp. 46–61, 2014, doi: 10.1016/j.advengsoft.2013.12.007.
[31] LendingClub. [Online]. Available: https://www.kaggle.com/competitions/ec-dsm-1-spring-2023-lendingclub/data.
[32] T. N. Truong, S. K. Thanh, T. N. T. Thu, N. N. Ha, and D. T. Manh, "Improve Risk Prediction in Online Lending (P2P) Using Feature Selection and Deep Learning," *International Journal of Computer Science and Network Security*, vol. 19, no. 11, pp. 216–222, 2019.
[33] L. Victor and M. Raheem, "Loan Default Prediction Using Genetic Algorithm: A Study within Peer-To-Peer Lending Communities," *International Journal of Innovative Science and Research Technology*, vol. 6, no. 3, pp. 1195–1205, 2021.

BIOGRAPHIES OF AUTHORS

Muhammad Sam'an received Bachelor Degree from Universitas Negeri Semarang and Master degree from Universitas Diponegoro in Mathematics 2010 and 2016 respectively. His research interests are in optimization, fuzzy mathematics, and computational mathematics. He can be contacted at email: muhammad92sam@unimus.ac.id.

Safuan received the Master degree in Informatics Engineering from Dian Nuswantoro University (UDINUS) in 2015. Currently, he is a lecturer in informatics Engineering at Muhammadiyah University, Semarang (UNIMUS). His research interests include data mining, programming, and web security. He can be contacted at email: safuan@unimus.ac.id.

Muhammad Munsarif received the Master degree in Computer science from Dian Nuswantoro University (UDINUS) in 2002. Currently, he is a lecturer in informatics Engineering at Muhammadiyah University, Semarang (UNIMUS). His research interests include computer vision, data science, and technopreneuership. He can be contacted at email: m.munsarif@unimus.ac.id.