Analysis of Prediction Model for Repayment Ability of Loan Funds Using XGBoost and Neural Network in Technical Internship Training Programs in Japan

Mohammad Roffi Suhendry, Gerry Firmansyah,
Nenden Siti Fatonah and Agung Mulyo Widodo

February 25, 2025

Analysis of Prediction Model for Repayment Ability of Loan Funds Using XGBoost and Neural Network in Technical Internship Training Programs in Japan

Mohammad Roffi Suhendry[1,] Gerry Firmansyah[2*,] Nenden Siti Fatonah[3,] , Agung Mulyo Widodo[4*]
[1,2,3,4]Universitas Esa Unggul
mohammad.roffi13@student.esaunggul.ac.id, gerry@esaunggul.ac.id, nenden.siti@esaunggul.ac.id, agung.mulyo@esaunggul.ac.id

Abstract

This study compares XGBoost and Multi-Layer Perceptron (MLP) models in predicting the delayed repayment of financial aid provided to technical internship training in Japan. The preprocessing stage includes handling missing data, normalization, and feature selection using a correlation threshold of 0.06, where features with an absolute correlation below this value are removed. The XGBoost Default, XGBoost with GridSearchCV, and MLP models are evaluated using accuracy, precision, recall, F1-score, and ROC-AUC metrics. The results show that XGBoost Default achieves the highest accuracy (95%) with precision of 95%, but a recall of only 83%, whereas XGBoost with GridSearchCV improves recall to 84% with a slight accuracy reduction (94%). The MLP model performs the worst (accuracy 92%, recall 74%), indicating difficulties in detecting delays. With a ROC-AUC of 91% for XGBoost compared to 86% for MLP, XGBoost proves to be the superior model. This model can assist training institutions providing financial aid programs in more accurately selecting participants, reducing repayment delays, and improving financial management efficiency.

Keywords: Machine Learning, XGBoost, Multi-Layer Perceptron, Late Payment Prediction, Loan Repayment.

1 Introduction

The Technical Internship Training Program was officially introduced in 1993 to transfer Japan's advanced skills and technology to developing countries, including Indonesia. This initiative aims to enhance international cooperation by fostering human resource development and supporting economic growth (JITCO, n.d.). Before departure, trainees must complete technical and language training at a Training Institution (LPK). However, they are required to bear educational and administrative costs, which often become a financial burden. Many trainees struggle to meet these obligations, highlighting the need for a more flexible financial support system (Hermann & Fauskanger, 2024).

To address this challenge, a private training institution offers a loan assistance program that covers training fees, administrative costs, and travel expenses. Trainees repay this loan within a year after securing employment in Japan (Guan, Suryanto, Mahidadia, Bain, & Compton, 2023). However, many trainees experience delayed repayments, causing financial strain on the institution. In 2022, 22% of

[*] Created the first draft of this document

trainees failed to meet payment deadlines, leading to Rp 4.5 billion in outstanding loans. This issue disrupts cash flow, hampers the institution's ability to fund new trainees, and threatens the sustainability of the program (Koc, Ugur, & Kestel, 2023).

This study proposes a machine learning model to predict repayment risks among prospective trainees. By implementing this predictive system, the institution can enhance its selection process, reduce financial risks, and ensure long-term program viability. Additionally, the model serves as a decision-support tool for managing loan assistance more efficiently, improving overall financial planning, and ensuring continued access to training opportunities for future trainees (Nurdin et al., 2023).

2 Literature Review

Several studies have investigated the use of machine learning models for credit risk classification, highlighting various optimization techniques and model performances. (Yulianti, Soesanto, & Sukmawaty, 2022) demonstrated that XGBoost achieved 80.02% accuracy in credit card classification, which improved to 83.42% after hyperparameter tuning, confirming that parameter optimization enhances model performance in detecting high-risk customers. Similarly, (Arram, Ayob, Albadr, Sulaiman, & Albashish, 2023) compared multiple models and found that MLP Neural Network performed best, achieving 91.6% accuracy with high recall, making it more effective for identifying default risks. However, while LightGBM attained the highest accuracy (94.4%), its lower recall made it less suitable for risk detection.

Meanwhile, (Rao, Liu, & Goh, 2023) focused on automotive credit risk and discovered that PSO-XGBoost outperformed other models, achieving 83.11% accuracy, 88.60% precision, and 76.14% recall, with an AUC-ROC of 0.90. The integration of PSO optimization and Smote-Tomek Link significantly improved accuracy and model stability, making it an ideal method for vehicle credit scoring. Additionally, (Li, Stasinakis, & Yeo, 2022) introduced a hybrid XGBoost-MLP model for credit risk assessment in digital supply chain finance, outperforming several traditional models, including Logistic Regression, KNN, Naïve Bayes, Decision Tree, Random Forest, and SVM. Their proposed model achieved 98.3% accuracy and an AUC of 0.994, proving to be the most effective method in detecting high-risk credit cases. The model's success is attributed to feature selection with XGBoost and the incorporation of digital supply chain features, enhancing its predictive capability.

Furthermore, (He et al., 2022) found that DNN-XGBoost demonstrated higher accuracy and stability in credit risk assessment, with AUC increasing by 8.6% and Type I error reduced by 54.2%. This model outperformed Logistic Regression, Random Forest, and SVM, proving to be an effective solution for banks and financial institutions in mitigating default risks. These findings reinforce the superiority of hybrid and ensemble learning approaches in credit risk prediction, offering more reliable and accurate tools for financial decision-making.

Machine learning has become essential in credit risk assessment, enabling financial institutions to make more accurate predictions. Among various models, XGBoost and Neural Networks have proven highly effective, outperforming traditional methods like Logistic Regression, Random Forest, and SVM. These advanced models offer better feature selection, model stability, and predictive accuracy.

Studies show that XGBoost-MLP achieves 98.3% accuracy with an AUC of 0.994, leveraging XGBoost's feature selection and MLP's deep learning strengths. Similarly, DNN-XGBoost improves AUC by 8.6% and reduces Type I error by 54.2%, enhancing risk detection. PSO-XGBoost also improves recall and model stability, making it ideal for financial risk assessment.

With growing demand for accurate and scalable credit evaluation models, hybrid approaches combining XGBoost and Neural Networks offer a powerful solution. This study explores their potential in optimizing credit risk prediction, ensuring higher accuracy and reliability.

3 Research Methods

3.1 Collection

In the data collection stage, relevant financial and recruitment data are selected to assess the repayment ability of technical internship training in Japan (Dwi et al., 2024). This study integrates financial transaction reports and recruitment data, including demographics and employment status, to identify key factors influencing loan repayment (Riyadi & Juliane, 2024).

3.2 Preprocessing

In the preprocessing stage, the collected data undergoes cleaning and preparation to ensure optimal use in the predictive model. This process is crucial to maintaining data quality, eliminating noise, and making it ready for further analysis (Riyadi & Juliane, 2024).

3.3 Model Classification

Model classification is a key stage in data mining, where machine learning is used to analyze patterns and build predictive models. At this stage, processed data is used to train and test models to predict the repayment ability of technical internship training in Japan. The results help identify key factors influencing repayment and improve credit evaluation efficiency (Riyadi & Juliane, 2024).

3.4 Evaluation

The evaluation stage measures the predictive model's performance using various metrics to ensure accuracy, stability, and real-world applicability. A well-evaluated model enhances the recruitment process for technical internship training in Japan by improving credit risk assessment and decision-making (Riyadi & Juliane, 2024).

4 Results and Discussion

4.1 Preprocessing

The data used in this study comes from financial reports and the recruitment process. The collected dataset includes various key features that support further analysis. Below is a list of features in the dataset, which will be utilized to extract insights and identify patterns in this research.

#	Column	Total Data	Data Type
1	jenis_kelamin	1882	object
2	jumlah_keluarga	1878	float64
3	umur_saat_berangkat	1882	Int64
4	usaha	1881	object
5	ketrampilan	1882	object

6	total_hutang	1881	Float64
7	tgl_tahun_pertama	1882	Datetime64
8	tgl_bayar_terakhir	1855	Datetime64
9	total_pembayaran	1855	Float64
10	komitmen	1882	Int64
11	izin_keluarga	1882	Int64
12	pelatihan_prakeberangkatan	1882	Int64
13	pekerjaan_dijepang	1882	Int64
14	kondisi_status_keluarga	1882	Int64
15	kesehatan	1882	Int64
16	sikap_kesopanan	1882	Int64
17	focus	1882	Int64
18	penilaian_keseluruhan	1880	Float64
19	hutang_pribadi	1882	Int64
20	hutang_orang_tua	1882	Int64
21	telat_bayar	1882	Int64
22	iq	1750	Float64

Table 1: Features of dataset

The image presents a summary of a dataset consisting of 22 columns with various data types. The dataset contains a total of 1,882 rows, though some columns have fewer entries, indicating missing data. In the predictive modeling process, the telat_bayar column serves as the primary target variable to be predicted.

#	Count
0	1412
1	470

Table 2: Total Records of target

The image displays the data distribution in the 'telat_bayar' column, which serves as the target variable in the predictive model. The dataset consists of 1,412 entries labeled as 0, indicating on-time payments, and 470 entries labeled as 1, representing delayed payments.

Table 3: Display of features with null values

#	Column	Total Null Values
1	jenis_kelamin	0
2	jumlah_keluarga	4
3	umur_saat_berangkat	0
4	usaha	1
5	ketrampilan	0
6	total_hutang	1
7	tgl_tahun_pertama	0
8	tgl_bayar_terakhir	27
9	total_pembayaran	27
10	komitmen	0
11	izin_keluarga	0
12	pelatihan_prakeberangkatan	0
13	pekerjaan_dijepang	0
14	kondisi_status_keluarga	0
15	kesehatan	0
16	sikap_kesopanan	0
17	focus	0
18	penilaian_keseluruhan	2
19	hutang_pribadi	0
20	hutang_orang_tua	0
21	telat_bayar	0
22	iq	132

The missing values in the features **jumlah_keluarga, total_piutang, and penilaian_keseluruhan** are handled by filling them with the median to maintain data stability, especially when dealing with non-normal distributions or outliers. Meanwhile, the **usaha** feature is filled using the mode (most frequent value).

After handling missing values, categorical features in the dataset are processed using encoding techniques. Label Encoding is applied to convert categorical values into numerical representations for machine learning models. In this case, the features jenis_kelamin, usaha, and keterampilan are encoded.

Next, the researcher conducted feature engineering on the tgl_tahun_pertama and tgl_bayar_terakhir features to enhance data quality before modeling. Since both features are of the datetime type, a data transformation was applied to extract more relevant information. One of the steps taken was adding a new feature called selisih_hari_pembayaran, calculated by subtracting tgl_tahun_pertama from tgl_bayar_terakhir. This feature measures the time span between the last payment date and the initial date, providing deeper insights into payment delays or participant payment patterns. By incorporating this feature, the machine learning model can better recognize the relationship between payment timing and the likelihood of delays, thereby improving the accuracy of predicting repayment risks.

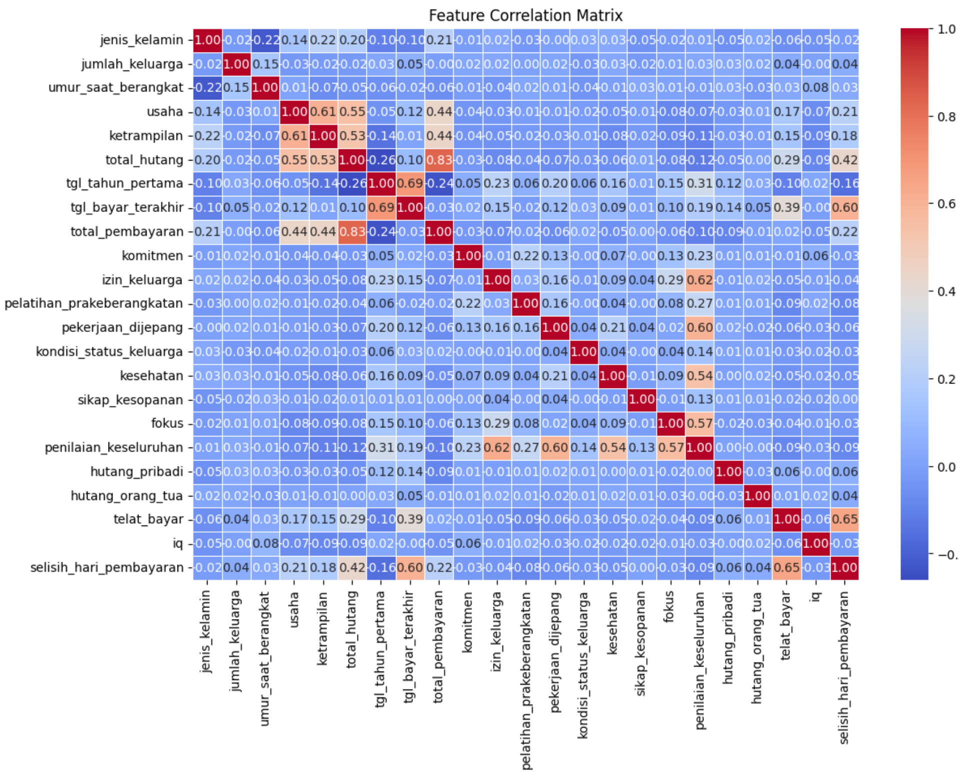

Figure 1: Feature Correlation Matrix

Based on the feature correlation matrix above, two features, total_hutang and total_pembayaran, exhibit a very high correlation. To prevent redundancy and multicollinearity issues, one of these features must be removed from the dataset. High correlation between two features can lead to overfitting and reduce the interpretability of the model, making feature selection an essential step in data preprocessing.

Additionally, any feature with a correlation of 0.06 or lower with telat_bayar will also be removed. A very low correlation indicates that the feature has little to no influence on the target variable, making it unnecessary for prediction. Removing such features helps streamline the dataset by eliminating irrelevant variables, thereby enhancing computational efficiency and model performance.

By removing redundant and weakly correlated features, we can improve model efficiency, reduce dimensionality, and minimize noise that does not contribute significantly to predictive accuracy. This process ensures that the model focuses only on the most relevant features, ultimately leading to better generalization and more accurate predictions. If additional features are found to have very low or near-zero correlation, further feature selection should be considered to enhance the model's reliability and robustness.

4.2 Model Classification

The dataset is split into 80% for training and 20% for testing, followed by the training process. After training, predictions are performed on the test data.

Algorithm	Accuracy	Precision	Recall	F1-Score	ROC-AUC
XGBoost Default	95%	95%	83%	89%	91%
XGBoost with GridSearchCV	94%	93%	84%	88%	91%
MLP Neural Network	92%	93%	74%	83%	86%

Table 4: Comparation Results

The evaluation results indicate that XGBoost Default achieves the highest accuracy (95%), making it the most reliable model for overall predictions. However, XGBoost with GridSearchCV demonstrates higher recall (84%) compared to XGBoost Default (83%), meaning it is better at identifying delayed payments, even though its accuracy is slightly lower at 94%. On the other hand, the MLP Neural Network performs the worst, with an accuracy of 92% and a recall of only 74%, indicating that it frequently fails to detect delayed repayment cases. Additionally, ROC-AUC scores show that XGBoost (91%) outperforms MLP (86%), proving that XGBoost is more effective in distinguishing between on-time and delayed payments. Overall, XGBoost is the superior model for predicting repayment risks, and the GridSearchCV optimization enhances recall, making it a more balanced choice for financial risk assessment in loan assistance programs.

4.3 Evaluation

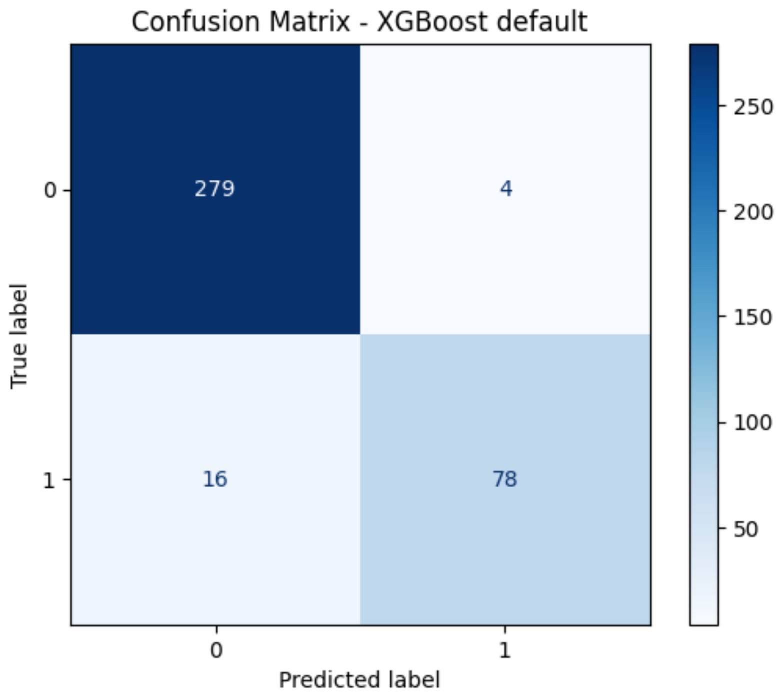

Figure 2: Confusion Matrix XGBoost Default

The confusion matrix for the XGBoost Default model demonstrates its performance in a binary classification task, where it successfully classifies most cases with high accuracy (94.7%). The model correctly identifies 279 true negative cases (non-defaulters) and 78 true positive cases (delayed payments). However, it misclassifies 4 cases as false positives, incorrectly labeling non-defaulters as defaulters, and fails to detect 16 actual defaulters (false negatives). With a precision of 95.1%, the model is highly reliable when predicting delayed payments, as 95.1% of its positive predictions are correct. Meanwhile, its recall of 83% indicates that the model successfully identifies 83% of actual defaulters but misses the remaining 17%, suggesting a slight trade-off between precision and recall.

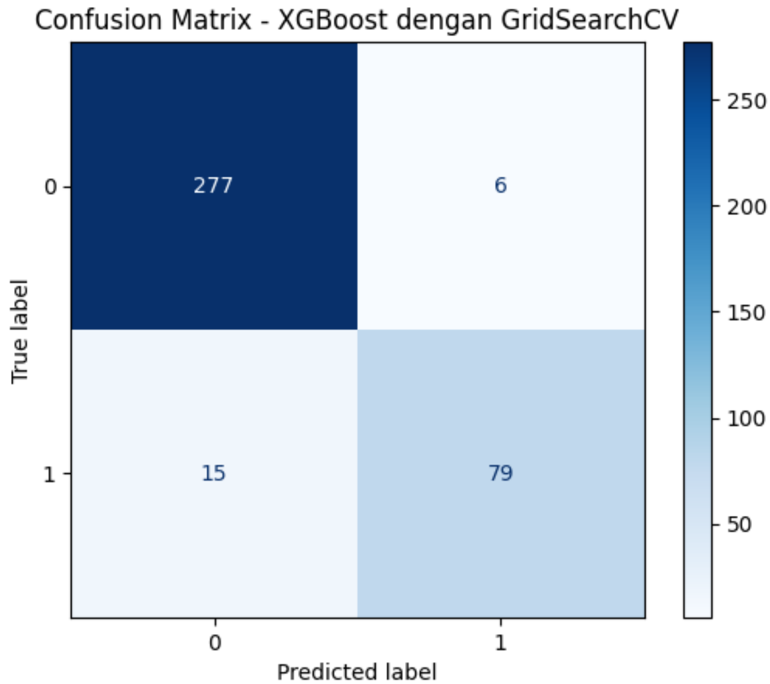

Figure 3: Confusion Matrix XGBoost With GridSearchCV

The confusion matrix for the XGBoost model optimized with GridSearchCV shows high accuracy in binary classification. The model correctly identifies 277 true negatives (TN) and 79 true positives (TP), while misclassifying 6 false positives (FP) and 15 false negatives (FN). Compared to the default XGBoost model, this version improves recall, enhancing its ability to detect delayed payments. However, the slight increase in false positives indicates a minor trade-off. Overall, the optimized model offers a better balance between precision and recall, making it more effective for predicting repayment risks.

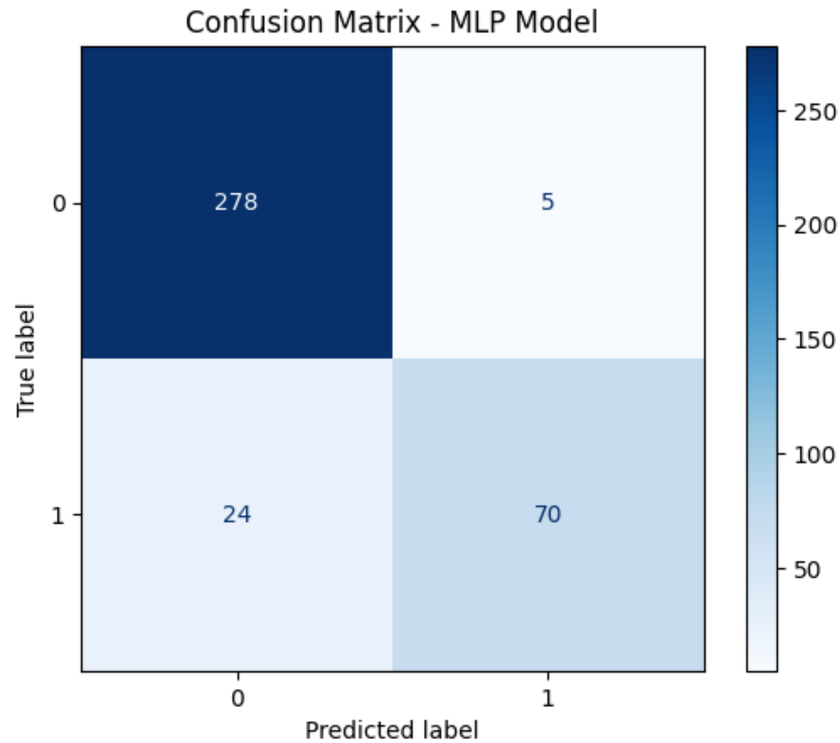

Figure 4: Confusion Matrix MLP

The confusion matrix for the MLP model shows that it correctly classifies 278 true negatives (TN) and 70 true positives (TP) but misclassifies 5 false positives (FP) and 24 false negatives (FN). Compared to XGBoost, MLP has lower recall, missing more delayed payments. While maintaining high accuracy, its weaker ability to detect repayment delays makes it less effective for financial risk assessment.

5 Conclusion

This study evaluates the performance of XGBoost and MLP (Multi-Layer Perceptron) in predicting repayment delays for loan assistance in technical internship training programs. The results show that XGBoost Default achieves the highest accuracy (95%), making it the most reliable model for overall prediction. However, XGBoost with GridSearchCV improves recall (84%) compared to XGBoost Default (83%), showing a better ability to detect delayed repayments while maintaining a high accuracy of 94%. On the other hand, the MLP model underperforms, with an accuracy of 92% and recall of only 74%, indicating its difficulty in detecting delayed payments. Additionally, XGBoost achieves a higher ROC-AUC score (91%) compared to MLP (86%), proving its superiority in distinguishing between timely and delayed payments. These findings suggest that XGBoost is the best model for predicting repayment risks, and hyperparameter tuning with GridSearchCV can enhance recall, making the model more balanced for financial risk assessment.

To further improve predictive performance, future research should explore feature engineering enhancements, such as time-series-based features, to better capture financial behavior. Additionally, alternative optimization techniques like Bayesian Optimization or Genetic Algorithms can be considered for more efficient hyperparameter tuning. Addressing class imbalance through SMOTE

(Synthetic Minority Over-sampling Technique) or cost-sensitive learning may improve recall. Furthermore, incorporating explainability techniques such as SHAP or LIME would enhance model interpretability, making it more transparent for financial institutions. Finally, deploying and testing the model in real-world scenarios could validate its effectiveness and refine it for practical use. By implementing these improvements, future research can develop more robust and reliable models for predicting repayment risks and financial decision-making

References

Arram, A., Ayob, M., Albadr, M. A. A., Sulaiman, A., & Albashish, D. (2023). *Credit card score prediction using machine learning models: A new dataset*. Retrieved from http://arxiv.org/abs/2310.02956

Dwi, E., Aini, N., Khasanah, R. A., Ristyawan, A., Diniati, E., Nusantara, U., & Kediri, P. (2024). Penggunaan Data Mining untuk Prediksi tingkat Obesitas di Meksiko Menggunakan Metode Random Forest. In *Agustus* (Vol. 8). Online.

Guan, C., Suryanto, H., Mahidadia, A., Bain, M., & Compton, P. (2023). Responsible Credit Risk Assessment with Machine Learning and Knowledge Acquisition. *Human-Centric Intelligent Systems*, *3*(3), 232–243. https://doi.org/10.1007/s44230-023-00035-1

He, X., Li, S., He, X. T., Wang, W., Zhang, X., & Wang, B. (2022). A Novel Ensemble Learning Model Combined XGBoost With Deep Neural Network for Credit Scoring. *Journal of Information Technology Research*, *15*(1), 1–18. https://doi.org/10.4018/jitr.299924

Hermann, R. R., & Fauskanger, E. A. (2024). Institutionalizing international internships in business education: an action research approach to overcoming barriers and driving systemic change in Norwegian business schools. *Scandinavian Journal of Educational Research*. https://doi.org/10.1080/00313831.2024.2369882

JITCO. (n.d.). What is the Technical Intern Training Program? Retrieved December 22, 2024, from https://www.jitco.or.jp website: https://www.jitco.or.jp/en/regulation/

Koc, O., Ugur, O., & Kestel, A. S. (2023). *The Impact of Feature Selection and Transformation on Machine Learning Methods in Determining the Credit Scoring*. Retrieved from http://arxiv.org/abs/2303.05427

Li, Y., Stasinakis, C., & Yeo, W. M. (2022). A Hybrid XGBoost-MLP Model for Credit Risk Assessment on Digital Supply Chain Finance. *Forecasting*, *4*(1), 184–207. https://doi.org/10.3390/forecast4010011

Nurdin, A., Zunaidi, R. A., Arkan, M., Wicaksono, F., Lobita, A., & Martadinata, J. (2023). *ANALISIS KREDIT PEMBAYARAN BIAYA KULIAH DENGAN PENDEKATAN PEMBELAJARAN MESIN*. *10*(2), 271–280. https://doi.org/10.25126/jtiik.2023106301

Rao, C., Liu, Y., & Goh, M. (2023). Credit risk assessment mechanism of personal auto loan based on PSO-XGBoost Model. *Complex and Intelligent Systems*, *9*(2), 1391–1414. https://doi.org/10.1007/s40747-022-00854-y

Riyadi, T., & Juliane, C. (2024). Antisipasi Persedian Oli Sepeda Motor Wilayah Karawang dengan Menggunakan Pendekatan Knowledge Discovery in Database (KDD)id 2 1,2 Sekolah Tinggi Manajemen Informatika dan Komputer LIKMI Bandung. *The Indonesian Journal of Computer Science Www.Ijcs.Net*, *13*(5). https://doi.org/10.33022/ijcs.v13i5.3992

Yulianti, E. H., Soesanto, O., & Sukmawaty, Y. (2022). Penerapan Metode Extreme Gradient Boosting (XGBOOST) pada Klasifikasi Nasabah Kartu Kredit. *JOMTA Journal of Mathematics: Theory and Applications*, *4*(1).