

Comparison of Random Forest and SVM Algorithms in Credit Risk Evaluation Based on Debtor Occupation

Putri Armilia Prayesy¹ and Angga Pujakesuma²

¹Software Engineering Technology, Department of Informatics and Business, Politeknik Manufaktur Negeri Bangka Belitung, Bangka, Indonesia

²Retail Management, Faculty of Economics and Humanities, Institut Teknologi dan Bisnis Nasional, Banyuasin, Indonesia

Corresponding author: Putri Armilia Prayesy (e-mail: putri@polman-babel.ac.id).

ABSTRACT Credit is one of the main sources of income for banking institutions and plays a crucial role in supporting long-term profit growth. However, credit distribution is inherently associated with risks, especially the risk of default when borrowers fail to meet their repayment obligations as agreed. One effective strategy to minimize such risks is to conduct a comprehensive and accurate creditworthiness assessment of prospective borrowers before loan approval is granted. This study aims to evaluate the performance of three classification algorithms—Random Forest, Support Vector Machine (SVM), and Artificial Neural Network (ANN)—in predicting credit risk based on the borrower's occupation. The dataset used consists of 1,314 loan records with an imbalanced distribution between performing and non-performing loans. The experimental results show that the Random Forest algorithm achieved the highest accuracy at 97%, followed by Support Vector Machine at 73% and Artificial Neural Networks at 64%. While ANN is capable of capturing complex patterns through multilayered learning, Random Forest proved to be the most effective and robust in handling the given dataset. These findings clearly indicate that Random Forest can serve as a reliable method for financial institutions to enhance credit risk evaluation and minimize potential losses arising from loan defaults.

KEYWORDS Credit Risk Evaluation, Credit Debtors, Random Forest, Support Vector Machine

I. INTRODUCTION

A bank is financial service institutions that collect funds from customers in the form of deposits and carry out operational activities to serve customers through deposit and withdrawal transactions, investments, customer complaint services, and the distribution of loans to customers [1]. This also includes cash management, payment processing, foreign exchange services, vehicle and housing credit financing, as well as providing credit solutions for companies, retail entrepreneurs, and employees [2].

One of the risks faced by banks when granting loans to prospective debtors is the borrower's failure to make installment payments on time or the tendency to delay payments, which can lead to non-performing loans (NPLs) [3]. To avoid such bad loans, banks establish specific requirements during the loan application process as part of their credit risk mitigation measures. To address and reduce the number of problematic loans, computational algorithms supported by intelligent systems are needed to assist banks in selecting eligible debtors.

This study analyzes the risk of loan distribution from various aspects considered in the process of approving individual loans, based on several criteria established as banking standards. The researchers employed random forest, support vector machine, and Artificial Neural Networks algorithms to compare predictive accuracy in loan repayment classification, using Python as the supporting tool. Technically, models such as Random Forest offer good interpretability through feature importance analysis, which can reveal dominant variables such as payment history, debt-to-income ratio, or job stability in influencing credit decisions [4]. Meanwhile, Support Vector Machine (SVM) enables analysis of support vectors to understand decision boundaries, although its clarity is more limited compared to tree-based models. Conversely, Artificial Neural Networks (ANN) have a much more complex structure, making interpretation more challenging. To overcome this limitation, additional interpretability methods such as Local Interpretable Model-agnostic Explanations (LIME) or SHapley Additive exPlanations (SHAP) can be employed to

explain each feature's contribution to the prediction outcome.

In the study by Muryono [5], the results of accuracy from the K-Nearest Neighbor (K-NN), Decision Tree, and Naïve Bayes algorithms were compared. The algorithm with the highest accuracy was then applied to determine creditworthiness. This research used 11 attributes, and through evaluation and validation with the 5-fold cross-validation method using RapidMiner, the highest accuracy was achieved by the Decision Tree (C4.5) algorithm, reaching 98% in the third test.

Andriani [6] conducted research comparing the performance of four classification algorithms: Support Vector Machine, Naïve Bayes, Random Forest, and Decision Tree. Based on accuracy, precision, recall, F1 score, and AUC-ROC metrics, the Decision Tree achieved the best performance with 42.5% accuracy, 48.3% precision, 47.5% recall, 47.5% F1 score, and an AUC of 0.60, indicating a moderate ability to distinguish creditworthiness. The study recommends implementing the Decision Tree algorithm with optimization through hyperparameter tuning, adding relevant features, and addressing data imbalance.

Hazizah & Feranika [7] investigated the implementation of the Random Forest algorithm in classifying the risk of credit card default among bank customers, showing that the algorithm achieved 81% accuracy in assessing default risk. The model's performance was significantly influenced by key criteria such as previous payment history, credit limits, and total bill amounts.

Although the model demonstrated good performance for non-defaulting customers, challenges remain in classifying defaulting customers, especially due to data imbalance. In addition to providing a basis for model improvement, this research demonstrates that Random Forest is valuable in supporting decision-making in the banking sector. Each algorithm has different characteristics, requiring appropriate hyperparameter tuning to improve performance—such as kernel and regularization parameters for Support Vector Machine, number of trees and maximum depth for Random Forest, as well as the number of neurons and learning rate for Artificial Neural Network.

In practice, data on prospective borrowers is often imbalanced, with the number of defaulters being much smaller than those who repay on time. This can reduce predictive accuracy for the minority class. To address this, handling techniques such as Synthetic Minority Oversampling Technique (SMOTE), undersampling, or cost-sensitive learning can be applied so that the model is not biased toward the majority class. From an interpretability perspective, Random Forest and Support Vector Machine can provide insights into feature importance or support vector analysis, while Artificial Neural Network tends to be more complex and requires additional interpretability methods such as lime or shap to explain predictions. This interpretability is crucial as it enables risk analysts and bank management to understand the reasoning behind the system's decisions. Practically, implementing an accurate credit

scoring model can help banks select borrowers more efficiently, reduce credit risk, and improve loan portfolio quality.

With the support of machine learning technology and appropriate optimization techniques, such a system can be integrated into real-time automated credit decision-making processes, adding value in terms of both operational efficiency and banking risk management. From the various studies reviewed, it is clear that there are differences in the performance of algorithms used to determine loan eligibility and their resulting accuracy levels. The aim of this research is to analyze the classification process and accuracy outcomes from comparing the Support Vector Machine (SVM), Random Forest, and Artificial Neural Network (ANN) methods to identify the best model for this study. Thus, while deep learning-based models often offer higher accuracy, combining them with interpretability techniques will ensure that the application of deep learning in banking remains ethical, transparent, and sustainable.

II. METHODOLOGY

A. DATASET AND PREPROCESSING

In this study, the authors will compare credit risk assessment using three machine learning algorithms: Random Forest, Support Vector Machine (SVM), and Artificial Neural Network (ANN). The selection of these three methods is based on their respective characteristics and capabilities in classifying data, particularly in the context of credit repayment performance of debtors. Support Vector Machine is known for its effectiveness in distinguishing classes with an optimal separating margin, Random Forest excels in interpretability and its ability to handle complex data, while Artificial Neural Network offers the potential for high accuracy despite its more challenging interpretability [8].

In classifying the credit repayment performance of debtors, several variables are used to evaluate payment accuracy, including occupation type, credit application limit, loan term (tenor), credit installment amount, income level, length of employment, number of dependents, payroll system type, and credit status as the target variable.

To ensure the research remains focused and aligned with the intended objectives, the study adopts a structured research framework [9]. The methodology includes stages of data collection, data preprocessing, splitting the dataset into training and testing sets, training models using the three selected algorithms, and evaluating model performance based on evaluation metrics such as accuracy, precision, recall, and F1-score. The flowchart illustrating the overall research process can be seen in Figure 1.

The flowchart of the comparison of Random Forest, Support Vector Machine and Artificial Neural Network Algorithms in Credit Risk Evaluation Based on Debtor Profession is shown in Figure 2.

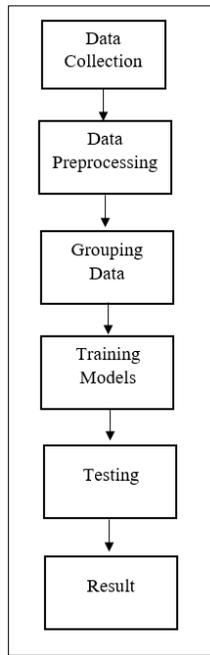


Figure 1. Research Flowchart

B. DATA RESEARCH

To support this research, data collection methods are needed. The research data in Table I consists of sources, data collection techniques, types, and sources of data along with data analysis.

The source of the data used in this research is based on data related to credit given to debtors, while the source of data used is secondary data. secondary data is something on existing data.

C. ALGORITHM ANALYSIS

1) SUPPORT VECTOR MACHINE

This algorithm is one of the supervised learning methods. Compared to other classification techniques, SVM has a more well-established mathematical concept, allowing it to handle both linear and non-linear classification problems [10]. for the Support Vector Machine (SVM), the kernel used was the Radial Basis Function (RBF), as it is capable of capturing non-linear patterns in the data. The regularization parameter was set to $C = 1.0$ to ensure that the separating margin was neither too tight nor too loose. The value of gamma was set to *scale* to adjust according to the number of features used. Similar to Random Forest, the class_weight parameter was also set to *balanced* to handle imbalanced class distribution. Additionally, output probability (probability = True) was enabled so that prediction results could be utilized in the ensemble method.

2) RANDOM FOREST

Random Forest is an algorithm built from multiple decision trees and is essentially a supervised learning method. It is a type of technique that can be used for classification and regression.

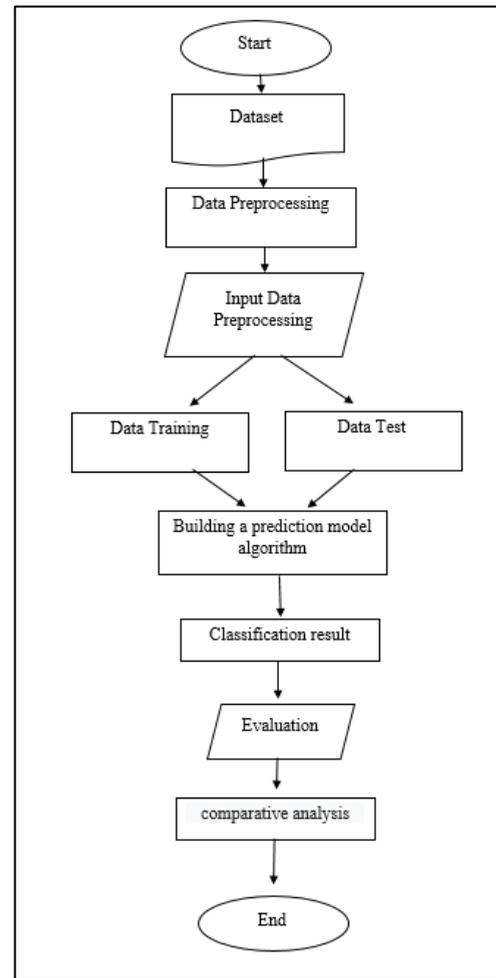


Figure 2. Flowchart Classification

The advantages of Random Forest lie in its ability to handle large data sets with numerous features and its ability to provide feature importance estimates that are useful for model interpretability [11]. In the context of credit risk assessment, feature importance analysis can help banks identify the variables that most influence a debtor's repayment schedule, such as debt-to-income ratio, history of late payments, or monthly installment amounts. However, Random Forest performance can decline if the data has a high class imbalance, requiring the application of preprocessing techniques such as SMOTE or class weight adjustment to maintain accuracy in minority classes. In the Random Forest model, the number of trees was set to 200. This decision was based on the trade-off between accuracy and computational time: the more trees used, the more stable and accurate the predictions become, as the voting results are more representative. However, using too many trees increases computational time without providing significant performance improvements. The tree depth was left unrestricted, allowing the model to learn more complex data patterns. The minimum number of samples required to split a node was set to two, and the minimum number of samples for a leaf node was set to one. The number of features

TABLE I
 SAMPLE DATA

Job	Limit	Time Period	Installment	Income	Length of Work	Dependents	Payroll	Collateral	Credit Status
BUMN	Rp. 800.000.000	96	Rp. 1,957,430	> 25	> 5	2	Payroll	Yes	Good
BUMN	Rp. 300.000.000	72	Rp. 1,652,042	> 25	> 5	3	Payroll	Yes	Good
SWASTA	Rp. 100.000.000	60	Rp. 1,013,682	> 25	> 5	2	Payroll	Yes	Good
SWASTA	Rp. 200.000.000	84	Rp. 1,143,654	> 25	> 5	3	Payroll	Yes	Good
BUMN	Rp. 500.000.000	72	Rp. 1,480,442	> 25	> 5	3	Payroll	Yes	Good
SWASTA	Rp. 200.000.000	48	Rp. 1,299,800	> 25	> 5	2	Payroll	Yes	Good
SWASTA	Rp. 300.000.000	60	Rp. 1,045,681	> 25	> 5	1	Payroll	Yes	Good
SWASTA	Rp. 100.000.000	60	Rp. 1,026,730	> 25	> 5	1	Payroll	Yes	Good
PNS	Rp. 70.000.000	60	Rp. 871,080	> 25	> 5	1	Payroll	Yes	Good
BUMN	Rp. 100.000.000	36	Rp. 3,013,376	> 25	> 5	3	Payroll	Yes	Good
BUMN	Rp. 300.000.000	84	Rp. 2,637,527	> 25	> 5	3	Payroll	No	Arrear
SWASTA	Rp. 300.000.000	60	Rp. 2,640,747	> 25	> 5	3	Payroll	No	Arrear
SWASTA	Rp. 100.000.000	60	Rp. 1,161,102	> 25	> 5	1	Payroll	No	Good
SWASTA	Rp. 500.000.000	84	Rp. 1,564,663	> 25	> 5	2	Payroll	No	Good
BUMN	Rp. 200.000.000	72	Rp. 1,319,682	> 25	> 5	2	Payroll	No	Good
SWASTA	Rp. 70.000.000	48	Rp. 1,193,375	> 25	> 5	1	Payroll	Yes	Good
SWASTA	Rp. 100.000.000	60	Rp. 1,042,831	> 25	> 5	1	Payroll	Yes	Good
SWASTA	Rp. 200.000.000	60	Rp. 1,323,850	> 25	> 5	2	Payroll	Yes	Good
SWASTA	Rp. 100.000.000	60	Rp. 1,284,653	> 25	> 5	2	Payroll	Yes	Good
SWASTA	Rp. 100.000.000	60	Rp. 1,184,812	> 25	> 5	1	Payroll	Yes	Good
BUMN	Rp. 70.000.000	36	Rp. 2,376,581	> 25	> 5	1	Payroll	Yes	Arrear

considered at each split was determined by the square root of the total number of features. To address class imbalance, the parameter `class_weight` was set to *balanced*, so that classes with fewer data points were given greater weight.

3) ARTIFICIAL NEURAL NETWORK

Artificial Neural Networks (ANN) are machine learning algorithms inspired by the functioning of biological neural networks in the human brain. ANNs consist of a collection of units called neurons, connected by weights and arranged in several layers: an input layer, a hidden layer, and an output layer [12]. The learning process is carried out by adjusting the connection weights between neurons using the backpropagation algorithm to minimize prediction errors. ANN are excellent at modeling complex nonlinear relationships, making them frequently used in various classification and prediction problems. However, compared to algorithms like SVM and Random Forest, ANN require larger data sets, require longer computation times, and present challenges in model interpretability.

In the context of credit risk evaluation, ANNs have proven effective in capturing complex relationships between variables such as income, credit history, and debt ratios, which may not be easily modeled by simple linear approaches. Artificial Neural Network (ANN) or Multi-Layer Perceptron (MLP), the network architecture consisted of two hidden layers, with 100 neurons in the first layer and 50 neurons in the second layer. The activation function used was ReLU (Rectified Linear Unit), as it accelerates the convergence process. Weight optimization was carried out using the Adam optimizer, with an initial learning rate of 0.001. Training was performed with a maximum of 500 iterations, and the early stopping method was applied to halt training earlier if no improvement in validation performance was observed.

4) EVALUATION

The final stage is to evaluate the level of success of the predictions made in data processing with the SVM and Random Forest methods so that the results of this study are useful for banking in order to reduce the level of bad credit [13]. The models that have been applied will be compared using the confusion matrix.

Confusion Matrix is one of the methods used to perform accurate calculations on the concept of data mining. The Confusion Matrix model will form a matrix consisting of true positive, true negative, false positive and false negative, as shown in (1), (2) and (3).

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

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

$$Precision = \frac{TP}{TP+FP} \quad (3)$$

Where:

- TP: True Positive
- TN: True Negative
- FP: False Positive
- FN: False Negative

III. RESULT AND DISCUSSION

The initial data in this study was processed by defining the training data, target data, and test data. The sample data was obtained from actual credit records consisting of 1,314 entries, including 1,089 performing credit records and 226 non-performing credit records, with 10 main attributes. Before the data mining process, a data cleaning stage was

carried out to remove duplicate entries, correct inconsistent data errors, and complete missing or incomplete data. All attributes in the dataset were selected after a relevance check to ensure that no attributes were redundant and all values were filled. Data was categorized as missing if an attribute contained no value or was blank, while data was considered redundant if the same record appeared more than once.

One of the main challenges in this dataset is the presence of class imbalance between the number of performing and non-performing credit records. This imbalance may cause the model to be more inclined to predict the majority class (performing credit), thus reducing its ability to detect the minority class (non-performing credit). To address this issue, several techniques for handling imbalanced data were applied, such as the Synthetic Minority Over-sampling Technique (SMOTE) to generate synthetic samples for the minority class, Random Under-Sampling to reduce the number of majority class records, and class weight adjustment in the learning algorithms to assign higher penalties for misclassification of the minority class.

After the preprocessing and data balancing stages were completed, the dataset was processed using three main algorithms: Support Vector Machine (SVM), Random Forest, and Artificial Neural Network (ANN). The classification process began by grouping data according to relevant variables, followed by training the models on the training set and testing them on the test set. The evaluation results, using metrics such as accuracy, precision, recall, F1-score, and AUC-ROC, provide an overview of each model's performance in predicting credit payment quality. This information can be utilized by the bank as a strategic consideration in making credit approval decisions for prospective borrowers.

A. SUPPORT VECTOR MACHINE

Confusion Matrix :				
Classification Report :				
	precision	recall	f1-score	support
Lancar	0.96	1.00	0.98	1088
Macet	1.00	0.80	0.89	226
accuracy			0.97	1314
macro avg	0.98	0.90	0.93	1314
weighted avg	0.97	0.97	0.96	1314

Figure 3. Results of Confusion Matrix of Support Vector Machine

The accuracy obtained from the SVM algorithm (in Figure 3) using 10 attributes and implemented in Python tools resulted in an accuracy of 73%, a class recall of 74%, and an F1-score for the *Lancar* (smooth) prediction of 65%, with a data support of 1314. The model predicted 783 true positives (TP), 305 false negatives (FN), 55 false positives (FP), and 171 true negatives (TN).

B. RANDOM FOREST

Confusion Matrix (Full Data):				
Classification Report				
	precision	recall	f1-score	support
Lancar	0.93	0.72	0.81	1088
Macet	0.36	0.76	0.49	226
accuracy			0.73	1314
macro avg	0.65	0.74	0.65	1314
weighted avg	0.84	0.73	0.76	1314

Figure 4. Results of Confusion Matrix of Random Forest

The accuracy obtained from the Random Forest algorithm (in Figure 4) using 10 attributes and implemented with Python tools resulted in an accuracy of 97%, a class recall of 90%, and an F1-score for the *Lancar* (smooth) prediction of 93%. With a data support of 1314, the model predicted 1088 true positives (TP), 0 false negatives (FN), 45 false positives (FP), and 181 true negatives (TN).

C. ARTIFICIAL NEURAL NETWORK

Confusion Matrix :				
Classification Report :				
	precision	recall	f1-score	support
Lancar	0.88	0.65	0.75	1088
Macet	0.25	0.55	0.34	226
accuracy			0.64	1314
macro avg	0.56	0.60	0.55	1314
weighted avg	0.77	0.64	0.68	1314

Figure 5. Results of Confusion Matrix of Artificial Neural Network

The accuracy in Figure 5 obtained from the *Artificial Neural Network* algorithm using 10 attributes and implemented with Python tools resulted in an accuracy of 64%, a class recall of 60%, and an F1-score for the *Lancar* (smooth) prediction of 55%. With a data support of 1314, the model predicted 711 true positives (TP), 377 false negatives (FN), 101 false positives (FP), and 125 true negatives (TN).

D. ENSEMBLE METHOD

Confusion Matrix Ensemble (Total 1314 Data):				
Classification Report Ensemble (Total 1314 Data):				
	precision	recall	f1-score	support
Lancar	0.84	1.00	0.91	1088
Macet	0.95	0.09	0.17	226
accuracy			0.84	1314
macro avg	0.90	0.55	0.54	1314
weighted avg	0.86	0.84	0.79	1314

Figure 6. Results of Confusion Matrix of Ensemble Method

In Figure 6, accuracy obtained from the Ensemble method using weighted voting based on confidence scores random forest, SVM, and ANN algorithms with predetermined attributes and implemented in Python software resulted in 84% accuracy, 55% class recall, and a 54% F1 score for smooth prediction. With 1,314 datasets, the model predicted 1,087 true positives (TP), 1 false negative (FN), 205 false positives (FP), and 21 true negatives (TN).

E. PERFORMANCE COMPARISON BASED ON TEST RESULTS

After conducting testing using three algorithms, namely Support Vector Machine (SVM), Random Forest, and Artificial Neural Network the comparison table of the Confusion Matrix is shown in Table II:

TABLE II
PERFORMANCE COMPARISON

Algorithm	Accuracy	Recall	F1-Score
Support Vector Machine (SVM)	73%	74%	65%
Random Forest	97%	90%	93%
Artificial Neural Networks (ANN)	64%	60%	55%
Ensemble Method	84%	55%	54%
Support Vector Machine (SVM)	73%	74%	

Based on the testing results, the Random Forest algorithm demonstrated better performance compared to the Support Vector Machine (SVM) and Artificial Neural Networks (ANN) algorithm. This is because Random Forest is an ensemble learning method consisting of multiple decision trees built randomly. The decision-making process in Random Forest is based on the voting results of each tree, making the model more stable and accurate, and better able to handle imbalanced data and noise. Random Forest is also capable of modeling non-linear relationships and attribute interactions, as each tree can learn from different subsets of features.

On the other hand, the SVM algorithm works by finding the optimal hyperplane that separates the data into two classes with the maximum margin. SVM is very effective in high-dimensional spaces and in cases where the number of features exceeds the number of samples. However, its performance may decline when dealing with non-linearly separable data or large and complex datasets, especially if kernel and parameter tuning are not properly performed.

Meanwhile, the Artificial Neural Network (ANN) mimics the way the human brain works by using interconnected layers of artificial neurons. ANN excels at learning complex and non-linear patterns, making it widely used for classification and prediction problems that are difficult to solve with traditional algorithms. However, ANN requires a large amount of data to achieve optimal performance and is quite sensitive to parameters such as the number of layers, number of neurons, and activation functions. In addition, ANN tends to require longer training times compared to

Random Forest and SVM. In this test, the performance of ANN was below that of Random Forest, likely due to the limited amount of data, potential overfitting, and the need for more complex parameter tuning.

Thus, it can be concluded that in this case, Random Forest outperformed SVM and ANN in terms of accuracy, recall, and F1-score, primarily because of its superior ability to handle varied data and noise.

IV. CONCLUSION

Based on the data mining process carried out in this study using the Knowledge Discovery in Database (KDD) approach, the Support Vector Machine (SVM) and Random Forest algorithms were implemented using Python tools to Comparison Random Forest and SVM Algorithms in Credit Risk Evaluation Based on Debtor Occupation.

From the test results, the Random Forest algorithm demonstrated superior performance. This is because Random Forest is an ensemble method consisting of multiple decision trees that operate collectively through a voting mechanism. This model is effective in handling complex and imbalanced data and is capable of identifying non-linear patterns by learning from different subsets of features.

On the other hand, the SVM algorithm works by finding the optimal hyperplane that separates classes with the maximum margin. SVM is highly effective for high-dimensional and well-structured data. However, its performance may decrease when the data is not linearly separable or when proper kernel and parameter tuning is not applied.

In conclusion, the Random Forest algorithm is more flexible and accurate in modeling complex patterns in the loan payment data, whereas the SVM algorithm is more suitable for structured and clearly separable datasets.

AUTHORS CONTRIBUTION

Putri Armilia Prayesy: drafting, conceptualization, methodology, coding, validation, Review Writing.

Angga Pujakesuma: Investigation, Writing & Editing Writing Draft.

COPYRIGHT



This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License.

REFERENCES

- [1] Purnamasidi, H., Darmawan Bakti, L., & Imran, B. (2022). SISTEM PAKAR PEMILIHAN JENIS KREDIT NASABAH MENGGUNAKAN METODE FORWARD CHAINING PADA PT. BANK RAKYAT INDONESIA (PERSERO). In *Jurnal Kecerdasan Buatan dan Teknologi Informasi (JKBTI)* (Vol. 1, Issue 3).
- [2] M. R. Givari, R. Mochamad, and Y. U. Sulaeman², "Perbandingan Algoritma SVM, Random Forest Dan XGBoost Untuk Penentuan Persetujuan Pengajuan Kredit," vol. 16, no. 1, 2022, [Online]. Available: <https://journal.uniku.ac.id/index.php/ilkom>

- [3] D. Wahyudi, P. Studi Manajemen Fakultas Ekonomi dan Bisnis Universitas Muhammadiyah Parepare Jl Jend Ahmad Yani, K. Parepare, and S. selatan, "ANALISIS KELAYAKAN PEMBERIAN KREDIT KUR PADA PT BANK RAKYAT INDONESIA CABANG PINRANG," vol. 2, no. 2, 2021, [online]. Available: <https://jurnal.umpar.ac.id/index.php/decision/article/view/1406/859>
- [4] Oktapiani, R., Prayudi, D., Fajria, A., Sella, N., Nufus, Z., & Lestari, R. N. (2022). Sistem Pendukung Keputusan Untuk Menentukan Manajemen Kelayakan Pemberian Kredit Di Bank Mandiri Taspen Sukabumi Menggunakan Metode Analytic Hierarchy Process. *Indonesian Journal on Software Engineering (IJSE)*, 8(1), 36–45. <http://ejournal.bsi.ac.id/ejurnal/index.php/ijse36>
- [5] T. T. Muryono, A. Taufik, and I. Irwansyah, "PERBANDINGAN ALGORITMA K-NEAREST NEIGHBOR, DECISION TREE, DAN NAIVE BAYES UNTUK MENENTUKAN KELAYAKAN PEMBERIAN KREDIT," *Infotech: Journal of Technology Information*, vol. 7, no. 1, pp. 35–40, Jun. 2021, doi: 10.37365/jti.v7i1.104.
- [6] W. Andriani, Gunawan, and N. N. P. W. Naja, "Analisis perbandingan machine learning untuk prediksi kelayakan kredit perbankan pada Bank BRI Tegal," *IT-Explore: Jurnal Penerapan Teknologi Informasi dan Komunikasi*, vol. 4, no. 1, pp. 82–92, Feb. 2025, doi: 10.24246/itexplore.v4i1.2025.pp82-92.
- [7] N. Hazizah and A. Feranika, "Implementasi Algoritma Random Forest Dalam Klasifikasi Risiko Gagal Bayar Kartu Kredit Pada Nasabah Bank," *Jurnal Manajemen Teknologi dan Sistem Informasi (JMS)*, vol. 5, no. 1, 2025, doi: 10.33998/jms.v5i1.
- [8] W. Wulansari and D. Purwitasari, "Algoritma Random Forest pada Prediksi Status Kredit Usaha Rakyat untuk Mengurangi Nonperforming Loan Rate," *Journal of Intelligent System and Computation*, vol. 5, no. 2, pp. 109–114, Oct. 2023, doi: 10.52985/insyst.v5i2.358.
- [9] Y. Susanti, R. Igo, A. Aziz, and M. Ahsan, "KLASIFIKASI KELAYAKAN PEMBERIAN KREDIT NASABAH BANK XYZ MENGGUNAKAN METODE ALGORITMA C4.5 DAN NAIVE BAYES." [Online]. Available: <https://jurnal.machung.ac.id/index.php/curawal>
- [10] B. Gaye, D. Zhang, and A. Wulamu, "Improvement of Support Vector Machine Algorithm in Big Data Background," *Math Probl Eng*, vol. 2021, 2021, doi: 10.1155/2021/5594899.
- [11] Gentleman, R., Hornik, K., & Parmigiani, G. (n.d.). "Random Forests with R Switzerland: Springer Nature," 2020. *Series Editors*. <http://www.springer.com/series/6991>
- [12] Martinelli, F., Mercaldo, F., Raucci, D., & Santone, A. (2020). Bank Credit Risk Management based on Data Mining Techniques. *International Conference on Information Systems Security and Privacy*, 837–843. <https://doi.org/10.5220/0009371808370843>
- [13] B. Prasajo and E. Haryatmi, "Analisa Prediksi Kelayakan Pemberian Kredit Pinjaman dengan Metode Random Forest," *Jurnal Nasional Teknologi dan Sistem Informasi*, vol. 7, no. 2, pp. 79–89, Sep. 2021, doi: 10.25077/teknosi.v7i2.2021.79-89.