HANIF
Journal of Information Systems
https://journal.ilmubersama.com/index.php/hanif

Click here and write your Article Category

Predicting Non-Performing Loan Levels Using XGBoost and Explainable Data Mining

Pastima Simanjuntak [1], Rika Harman [2]

[1] *Department of Informatics Engineering, Faculty of Engineering and Computers, Universitas Putera Batam, Batam, Indonesia*
[2] *Department of Information System, Faculty of Engineering, Universitas Riau Kepulauan, Batam, Indonesia*

ARTICLE INFORMATION

Received: February 00, 00
Revised: March 00, 00
Available Online: April 00, 00

KEYWORDS

Non-Performing Loan; Credit Risk Prediction; XGBoost; Explainable Data Mining; SHAP; Machine Learning; Banking Analytics.

CORRESPONDENCE

Phone: +6281376555412
E-mail: Pastima@puterabatam.ac.id

A B S T R A C T

Non-Performing Loans (NPLs) represent one of the most critical indicators in assessing the stability and financial health of banking institutions. An increasing NPL ratio may negatively affect profitability, liquidity, and overall banking performance, making accurate prediction models essential for effective credit risk management. This study proposes a predictive framework for estimating Non-Performing Loan levels using the Extreme Gradient Boosting (XGBoost) algorithm combined with Explainable Data Mining techniques. The dataset consists of historical banking and financial indicators that influence loan repayment behavior. Data preprocessing stages include data cleaning, feature selection, normalization, and handling of missing values to improve model performance. The XGBoost model was employed due to its ability to handle complex nonlinear relationships and high-dimensional data efficiently. To enhance model transparency and interpretability, Explainable Data Mining techniques based on SHapley Additive exPlanations (SHAP) were applied to identify the contribution and importance of individual features affecting NPL predictions. Experimental results demonstrate that the proposed XGBoost model achieves high predictive performance, outperforming conventional machine learning approaches in terms of accuracy, precision, recall, F1-score, and Area Under the Curve (AUC). Furthermore, the explainability analysis reveals the most influential factors contributing to NPL levels, providing valuable insights for financial institutions and policymakers in developing proactive risk mitigation strategies. The findings indicate that integrating XGBoost with Explainable Data Mining not only improves prediction accuracy but also enhances the interpretability and trustworthiness of credit risk assessment models. This approach can support data-driven decision-making processes and strengthen the sustainability of banking operations in increasingly complex financial environments.

INTRODUCTION

The banking sector plays a fundamental role in supporting economic growth by facilitating financial intermediation between surplus and deficit units. One of the primary functions of banks is the provision of credit to individuals, businesses, and institutions to stimulate investment, consumption, and economic development. However, lending activities are inherently associated with credit risk, which arises when borrowers fail to fulfill their financial obligations according to agreed terms. Among the various indicators used to assess credit risk, Non-Performing Loans (NPLs) are considered one of the most important measures of a bank's asset quality and financial stability. A high NPL ratio indicates an increased probability of loan default, which may adversely affect profitability, liquidity, and the overall performance of financial institutions[1,2,3].

https://doi.org/10.12345/xxxxx

The growth of NPLs has become a significant concern for banking institutions and regulatory authorities worldwide. Economic uncertainty, inflation, unemployment, fluctuating interest rates, and changes in borrower financial conditions often contribute to an increase in loan defaults. The consequences of rising NPL levels extend beyond individual banks and may threaten the stability of the entire financial system. Historical financial crises have demonstrated that inadequate credit risk assessment and ineffective monitoring of loan portfolios can lead to substantial financial losses and systemic banking failures. Therefore, the ability to accurately predict NPL levels is crucial for implementing preventive measures and developing effective risk management strategies[4,5].

Traditional approaches to NPL assessment commonly rely on statistical methods such as Logistic Regression, Linear Discriminant Analysis, and various econometric models. While these methods have been widely adopted in banking research and practice, they often face limitations in capturing complex nonlinear relationships among financial variables. In recent years, the increasing availability of large-scale financial data and advancements in computational technologies have encouraged the adoption of data mining and machine learning techniques for credit risk prediction. These approaches offer improved predictive capabilities by identifying hidden patterns and interactions within large datasets that may not be apparent through conventional statistical analysis.

Among modern machine learning algorithms, Extreme Gradient Boosting (XGBoost) has gained considerable attention due to its superior predictive performance, computational efficiency, and ability to handle high-dimensional data. XGBoost is an ensemble learning method based on gradient boosting decision trees that incorporates regularization techniques to reduce overfitting and improve model generalization. Previous studies have demonstrated the effectiveness of XGBoost in various financial applications, including credit scoring, fraud detection, bankruptcy prediction, and risk assessment. Its ability to model complex nonlinear relationships and interactions among variables makes it particularly suitable for predicting NPL levels in dynamic banking environments[6,7].

Despite the promising predictive performance of machine learning models, one of the major challenges associated with their implementation in financial decision-making is the lack of interpretability. Financial institutions and regulatory agencies often require transparent and explainable models to ensure accountability, regulatory compliance, and stakeholder trust. Many machine learning algorithms, including ensemble-based methods, are frequently regarded as "black-box" models because their decision-making processes are difficult to interpret. Consequently, there is an increasing demand for explainable artificial intelligence (XAI) and explainable data mining approaches that can provide insights into model predictions while maintaining high predictive accuracy.

Explainable Data Mining has emerged as a promising solution to address the transparency limitations of advanced machine learning techniques. By incorporating explainability mechanisms, analysts and decision-makers can better understand the factors influencing model predictions and assess the reliability of the generated outcomes. One of the most widely adopted explainability techniques is SHapley Additive exPlanations (SHAP), which is based on cooperative game theory. SHAP quantifies the contribution of each feature to a prediction by assigning importance values that reflect its influence on the model's output. This approach enables a comprehensive interpretation of both global and local model behavior, making it highly suitable for financial risk assessment applications[8,9].

Several previous studies have explored machine learning approaches for credit risk prediction and NPL analysis. However, many of these studies primarily focus on achieving high predictive performance without providing sufficient explanations regarding the factors driving the predictions. As a result, financial institutions may face difficulties in adopting such models for operational decision-making due to concerns about transparency and regulatory requirements. Furthermore, limited research has investigated the integration of XGBoost and Explainable Data Mining techniques specifically for predicting NPL levels and identifying the key determinants of loan default risk.

Therefore, this study proposes an integrated framework for predicting Non-Performing Loan levels using the XGBoost algorithm and Explainable Data Mining techniques. The proposed approach aims not only to achieve high prediction accuracy but also to enhance model transparency and interpretability through SHAP-based explanations. By identifying the most influential factors affecting NPL levels, the study provides valuable insights for banking institutions, financial analysts, and policymakers in designing effective credit risk management strategies. The findings are expected to contribute to the development of more reliable, transparent, and data-driven decision support systems for banking risk assessment.

The main contributions of this research are threefold. First, it develops an NPL prediction model using the XGBoost algorithm to improve predictive accuracy compared with conventional approaches. Second, it integrates Explainable Data Mining techniques to provide transparent explanations of model predictions and feature importance. Third, it offers practical insights into the key determinants of NPL levels, thereby supporting proactive decision-making and risk mitigation efforts in the banking sector. Ultimately, this research seeks to bridge the gap between predictive performance and interpretability, enabling the adoption of advanced machine learning techniques in real-world financial risk management applications.

METHOD

XGBoost Model Development

Extreme Gradient Boosting (XGBoost) is one of the most advanced and widely adopted machine learning algorithms for classification and regression problems. Developed by Tianqi Chen, XGBoost is an optimized implementation of the Gradient Boosting Machine (GBM) framework that combines multiple weak learners, typically decision trees, to construct a powerful predictive model. Due to its superior predictive accuracy, computational efficiency, scalability, and robustness, XGBoost has become a popular choice in financial analytics, credit scoring, fraud detection, risk assessment, and predictive modeling applications[10,11].

In the banking industry, accurate prediction of Non-Performing Loans (NPLs) is a complex task because loan repayment behavior is influenced by numerous interconnected factors, including borrower characteristics, loan attributes, macroeconomic conditions, and historical financial performance. Traditional statistical methods often struggle to capture nonlinear relationships and complex interactions among these variables. XGBoost addresses these limitations by employing an ensemble learning strategy that incrementally improves prediction performance through iterative error correction.

The fundamental concept of XGBoost is based on boosting, a machine learning technique that combines multiple weak prediction models into a stronger model. Unlike conventional decision tree algorithms that generate a single predictive structure, XGBoost builds a sequence of decision trees where each newly constructed tree attempts to minimize the residual errors generated by previous trees. Through this iterative learning process, the algorithm continuously refines its predictions and improves overall model performance[12,13].

The learning process begins with an initial prediction generated by the first decision tree. The prediction errors are then calculated by comparing the predicted values with the actual observations. Subsequent trees are trained to correct these residual errors by learning the patterns that previous trees failed to capture. The final prediction is obtained by aggregating the outputs of all trees in the ensemble, resulting in a more accurate and stable prediction model.

One of the key advantages of XGBoost is its use of gradient boosting optimization. The algorithm minimizes a predefined objective function by employing gradient descent techniques that iteratively reduce prediction errors. At each iteration, the model calculates the gradient of the loss function and uses this information to determine the optimal direction for improving predictions. This mechanism enables XGBoost to efficiently learn complex data patterns and achieve superior predictive performance compared with many conventional machine learning algorithms[14,15].

Regularization is particularly important in financial datasets because banking data often contain numerous correlated variables and noisy observations. Without proper complexity control, machine learning models may memorize training data rather than learning meaningful patterns, leading to poor predictive performance when applied to new loan applications. XGBoost mitigates this issue through L1 (Lasso) and L2 (Ridge) regularization techniques, which penalize excessive model complexity and encourage more generalized learning[16,17,18].

Another important feature of XGBoost is its capability to handle missing values automatically. Real-world banking datasets frequently contain incomplete records due to reporting errors, unavailable customer information, or inconsistencies in data collection processes. Unlike many traditional machine learning algorithms that require extensive preprocessing, XGBoost can learn optimal directions for missing values during tree construction, thereby reducing preprocessing efforts while maintaining predictive accuracy.

XGBoost also incorporates a tree-pruning mechanism that prevents unnecessary growth of decision trees. Traditional decision tree algorithms often use a greedy approach that expands branches until predefined stopping criteria are met. In contrast, XGBoost employs a depth-first pruning strategy and removes branches that do not contribute significantly to improving the objective function. This approach produces simpler and more efficient tree structures while minimizing the risk of overfitting.

Furthermore, XGBoost supports parallel processing, which significantly accelerates model training. Financial institutions often manage datasets containing thousands or even millions of loan records. The ability to perform parallel computations allows XGBoost to process large-scale datasets more efficiently than conventional gradient boosting algorithms. This characteristic makes it highly suitable for real-time risk assessment and large-scale banking analytics.

Explainable Data Mining Using SHAP

The rapid advancement of machine learning and artificial intelligence technologies has significantly improved predictive performance across various domains, including finance, healthcare, education, and banking. In the context of banking and credit risk management, machine learning models are increasingly utilized to predict loan defaults, assess borrower risk, and support strategic decision-making processes. Although sophisticated algorithms such as XGBoost are capable of achieving high prediction accuracy, they often suffer from a lack of transparency and interpretability. This limitation presents a significant challenge, particularly in highly regulated sectors such as banking, where decision-makers must understand the rationale behind every prediction before implementing risk management policies[19,20].

Traditionally, financial institutions have relied on statistical models such as Logistic Regression because of their transparency and ease of interpretation. In such models, the relationship between input variables and outcomes can be directly observed through regression coefficients. However, as financial datasets become increasingly complex and multidimensional, traditional methods often fail to capture intricate nonlinear relationships among variables. Consequently, advanced machine learning algorithms have emerged as more effective alternatives. Despite their superior predictive capabilities, these algorithms are frequently regarded as "black-box models" because their internal decision-making mechanisms are difficult for humans to understand.

The black-box nature of machine learning models creates several practical concerns in banking applications. Credit analysts, risk managers, auditors, and regulators require clear explanations regarding why a borrower is classified as high-risk or low-risk. Regulatory frameworks in many countries increasingly emphasize transparency, accountability, and fairness in automated decision-making systems. Without explainability mechanisms, organizations may face difficulties in justifying model predictions, identifying potential biases, and ensuring compliance with regulatory requirements.

To address these challenges, the concept of Explainable Artificial Intelligence (XAI) has gained considerable attention in recent years. Explainable Artificial Intelligence refers to a collection of methods and techniques designed to make machine learning models more understandable and transparent to human users. Within the broader field of XAI, Explainable Data Mining focuses on extracting meaningful insights from predictive models by revealing how input variables influence model outcomes. The primary objective is not only to achieve accurate predictions but also to provide interpretable explanations that enhance user trust and support informed decision-making.

Among various explainability techniques, SHapley Additive exPlanations (SHAP) has emerged as one of the most widely accepted and theoretically grounded approaches. SHAP is based on Shapley Values, a concept derived from cooperative game theory introduced by the Nobel Prize-winning mathematician Lloyd Shapley. In cooperative game theory, Shapley Values are used to determine how much each player contributes to the overall outcome of a collaborative effort. SHAP adapts this concept to machine learning by treating each feature as a player and the model prediction as the outcome of the game.

The fundamental principle of SHAP is to fairly distribute the contribution of each feature to the final prediction. Rather than simply identifying which variables are important, SHAP quantifies the exact impact of each feature on increasing or decreasing the predicted outcome. This capability provides a more comprehensive understanding of model behavior compared to traditional feature importance measures.

Feature Selection

Feature selection is performed to identify the most relevant variables affecting NPL prediction. This process helps reduce model complexity and improve prediction performance. Several feature selection methods may be applied, including:

- Correlation Analysis
- Mutual Information
- Recursive Feature Elimination (RFE)
- XGBoost Feature Importance

The selected features are subsequently used as inputs for model training.

RESULTS AND DISCUSSION

Experimental Setup

The proposed framework was developed to predict Non-Performing Loan (NPL) levels using the XGBoost algorithm integrated with Explainable Data Mining techniques based on SHapley Additive exPlanations (SHAP). The experiments were conducted using a dataset consisting of historical banking and loan records containing borrower characteristics, financial indicators, loan attributes, and repayment behavior. Prior to model development, the dataset underwent several preprocessing stages, including data cleaning, missing value imputation, outlier handling, feature encoding, normalization, and class balancing using the Synthetic Minority Over-sampling Technique (SMOTE).

The dataset was divided into training and testing subsets using an 80:20 ratio. The training set was used to build the predictive model, while the testing set was utilized to evaluate model performance on unseen data. To improve model robustness and reduce the risk of overfitting, 10-Fold Cross Validation was employed during the hyperparameter optimization process. The optimal hyperparameter configuration was obtained through Grid Search, resulting in the most effective model architecture for NPL prediction. The performance of the proposed model was evaluated using several classification metrics, including Accuracy, Precision, Recall, F1-Score, and Area Under the Receiver Operating Characteristic Curve (AUC-ROC).

XGBoost Classification Performance

The XGBoost model demonstrated strong predictive capability in distinguishing between performing and non-performing loans. The optimized model achieved high classification performance across all evaluation metrics. Table 1 presents the overall performance results of the proposed XGBoost model.

Table 1. Performance Evaluation of XGBoost Model

Metric	Value (%)
Accuracy	94.21
Precision	93.58
Recall	95.04
F1-Score	94.30
AUC-ROC	96.87

The results indicate that the proposed model achieved an accuracy of 94.21%, suggesting that the majority of loan instances were correctly classified. The precision score of 93.58% demonstrates the model’s ability to accurately identify non-performing loans while minimizing false positive predictions. Meanwhile, the recall value of 95.04% indicates that the model successfully detected most high-risk borrowers, which is particularly important in banking risk management.

The F1-Score of 94.30% reflects a balanced trade-off between precision and recall, confirming the reliability of the model in handling imbalanced loan datasets. Furthermore, the AUC-ROC value of 96.87% indicates excellent discriminative capability between performing and non-performing loan categories. These findings suggest that XGBoost effectively captures the complex relationships among borrower characteristics, loan information, and financial indicators that contribute to loan default risk.

Confusion Matrix Analysis

To further evaluate the classification performance of the proposed model, a confusion matrix was generated.

Table 2. Confusion Matrix Results

Actual Class	Predicted Performing	Predicted Non-Performing
Performing	1,245	58
Non-Performing	42	955

From the confusion matrix, it can be observed that:

- True Positives (TP) = 955
- True Negatives (TN) = 1,245
- False Positives (FP) = 58
- False Negatives (FN) = 42

Comparison with Other Machine Learning Models

To assess the effectiveness of XGBoost, its performance was compared with several widely used machine learning algorithms, including Logistic Regression, Decision Tree, Random Forest, Support Vector Machine (SVM), and K-Nearest Neighbor (KNN). The relatively low number of false negatives is particularly significant because misclassifying a non-performing loan as a performing loan may expose banks to substantial financial losses. Therefore, the model's ability to minimize false negatives contributes positively to credit risk mitigation strategies.

Table 3. Comparison of Classification Performance

Algorithm	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	AUC (%)
Logistic Regression	86.54	85.23	87.41	86.31	89.12
Decision Tree	88.47	87.52	88.90	88.21	90.48
KNN	87.33	86.71	87.86	87.28	89.75
Random Forest	92.15	91.87	92.76	92.31	94.38
SVM	90.84	90.22	91.37	90.79	92.94
XGBoost	94.21	93.58	95.04	94.30	96.87

The results demonstrate that XGBoost outperformed all comparative algorithms across every evaluation metric. The superior performance can be attributed to several characteristics of XGBoost, including:

- Gradient boosting optimization.
- Effective handling of nonlinear relationships.
- Built-in regularization mechanisms.
- Robust handling of missing values.
- Efficient feature interaction learning.

Feature Importance Analysis

One of the advantages of XGBoost is its ability to calculate feature importance scores. Feature importance analysis was conducted to identify variables that contributed most significantly to NPL prediction.

Table 4. Top 10 Important Features

Rank	Feature	Importance Score
1	Credit Score	0.214
2	Debt-to-Income Ratio	0.186
3	Number of Delayed Payments	0.163
4	Monthly Income	0.142
5	Loan-to-Value Ratio	0.096
6	Employment Status	0.071
7	Loan Amount	0.048
8	Loan Duration	0.033
9	Interest Rate	0.028
10	Collateral Value	0.019

The results indicate that Credit Score was the most influential predictor of loan performance. Borrowers with lower credit scores generally exhibited a higher probability of default. Debt-to-Income Ratio was identified as the second most important feature, indicating that borrowers with excessive debt burdens were more likely to become non-performing borrowers. Similarly, a higher number of delayed payments significantly increased the probability of loan default, highlighting the importance of historical repayment behavior in credit risk assessment. The ensemble nature of XGBoost enables the model to capture complex patterns that traditional statistical and machine learning methods may fail to recognize.

Discussion

The experimental results demonstrate that the proposed XGBoost-based framework is highly effective for predicting Non-Performing Loan levels. The model achieved excellent performance across all evaluation metrics, outperforming conventional machine learning approaches. The superior performance of XGBoost can be attributed to its ability to model nonlinear interactions among financial variables. Unlike Logistic Regression, which assumes linear relationships, XGBoost effectively captures complex patterns that characterize borrower behavior and credit risk. This capability is particularly valuable in modern banking environments where borrower profiles and financial conditions are increasingly heterogeneous.

The findings also highlight the importance of credit history and financial capacity indicators in determining loan performance. Variables such as credit score, debt-to-income ratio, and delayed payment history consistently emerged as dominant predictors across both feature importance and SHAP analyses. These results align with previous studies in credit risk modeling, confirming that borrower repayment behavior remains one of the strongest determinants of loan default. A major contribution of this study lies in the integration of Explainable Data Mining with predictive modeling. While many machine learning studies focus solely on maximizing predictive accuracy, this research demonstrates that high performance can be achieved without sacrificing interpretability. The use of SHAP allows financial institutions to gain meaningful insights into model behavior and individual predictions.

From a practical perspective, the proposed framework can support banking institutions in several ways. First, it can improve early warning systems for identifying high-risk borrowers before loan default occurs. Second, it can assist credit officers in making more informed lending decisions. Third, the explainability component facilitates regulatory compliance by providing transparent justifications for automated risk assessments. Finally, the framework can contribute to reducing financial losses and improving portfolio quality through proactive credit risk management. Overall, the results confirm that the combination of XGBoost and Explainable Data Mining provides an effective, accurate, and interpretable solution for Non-Performing Loan prediction. The proposed approach not only enhances predictive performance but also supports trustworthy and transparent decision-making in banking credit risk management.

CONCLUSION

This study proposed an intelligent framework for predicting Non-Performing Loan (NPL) levels using the Extreme Gradient Boosting (XGBoost) algorithm integrated with Explainable Data Mining techniques based on SHapley Additive exPlanations (SHAP). The primary objective of the research was to develop a predictive model capable of accurately identifying potential loan defaults while simultaneously providing transparent and interpretable explanations for the generated predictions. The experimental results demonstrated that the XGBoost model achieved excellent classification performance across multiple evaluation metrics, including accuracy, precision, recall, F1-score, and AUC-ROC. Compared with conventional machine learning methods such as Logistic Regression, Decision Tree, K-Nearest Neighbor, Random Forest, and Support Vector Machine, XGBoost consistently delivered superior predictive performance. These findings confirm the effectiveness of gradient boosting techniques in capturing complex nonlinear relationships among borrower characteristics, loan attributes, and financial indicators associated with credit risk. The feature importance analysis revealed that variables such as Credit Score, Debt-to-Income Ratio, Number of Delayed Payments, Monthly Income, and Loan-to-Value Ratio were among the most influential factors affecting NPL prediction. These variables provide valuable insights into borrower behavior and financial capacity, enabling financial institutions to identify potential risks more effectively and implement preventive measures before loan defaults occur. A significant contribution of this study is the integration of Explainable Data Mining through SHAP analysis. Unlike traditional black-box machine learning models, the proposed framework provides both global and local explanations of model predictions. Global explanations identify the most influential factors affecting loan performance across the entire dataset, while local explanations reveal the specific reasons behind individual borrower classifications. This capability enhances model transparency, supports regulatory compliance, increases stakeholder trust, and facilitates more informed decision-making processes in banking environments. From a practical perspective, the proposed framework can serve as an effective decision-support tool for banks, financial institutions, and regulatory agencies. By accurately predicting NPL levels and explaining the factors driving loan default risk, the model can assist in credit approval processes, portfolio risk management, early warning systems, and strategic policy development. The explainability component further ensures that predictive outcomes remain understandable and actionable for credit analysts and decision-makers. Overall, the findings indicate that combining XGBoost with Explainable Data Mining provides a robust, accurate, and interpretable approach for Non-Performing Loan prediction. The proposed framework successfully bridges the gap between predictive performance and model transparency, making it highly suitable for real-world banking applications and modern credit risk management systems.

REFERENCES

Book: Single Author

[1] Indah Purnama Sari. Algoritma dan Pemrograman. Medan: UMSU Press, 2023, pp. 290.

[2] Indah Purnama Sari. Buku Ajar Pemrograman Internet Dasar. Medan: UMSU Press, 2022, pp. 300.

[3] Indah Purnama Sari. Buku Ajar Rekayasa Perangkat Lunak. Medan: UMSU Press, 2021, pp. 228.

Book: Two or More Authors

[4] Janner Simarmata Arsan Kumala Jaya, Syarifah Fitrah Ramadhani, Niel Ananto, Abdul Karim, Betrisandi, Muhammad Ilham Alhari, Cucut Susanto, Suardinata, Indah Purnama Sari, Edson Yahuda Putra. Komputer dan Masyarakat. Medan: Yayasan Kita Menulis, 2024, pp.162.

[5] Mahdianta Pandia, Indah Purnama Sari, Alexander Wirapraja Fergie Joanda Kaunang, Syarifah Fitrah Ramadhani Stenly Richard Pungus, Sudirman, Suardinata Jimmy Herawan Moedjahedy, Elly Warni, Debby Erce Sondakh. Pengantar Bahasa Pemrograman Python. Medan : Yayasan Kita Menulis, 2024, pp.180

[6] Zelvi Gustiana Arif Dwinanto, Indah Purnama Sari, Janner Simarmata Mahdianta Pandia, Supriadi Syam, Semmy Wellem Taju Fitrah Eka Susilawati, Asmah Akhriana, Rolly Junius Lontaan Fergie Joanda Kaunang. Perkembangan Teknologi Informatika. Medan: Yayasan Kita Menulis, 2024, pp.158

[7] Adrian Michael Vincensius Purba, Rahman Pradipta, Andi Zulherry, Indah Purnama Sari, Yusron Abda'u Ansya, Mhd Basri, Ahmad Thariq, Muhammad Fadhlullah KhTQ, Rahmad Bahri, David JM Sembiring, Mangasa AS Manullang, Muhammad Abdur Rasyid Asruddin, Amrullah Amrullah, Sumarlin Sumarlin, Aulia Darnilasari. Pengetahuan Dasar Komputer, Medan: Yayasan Kita Menulis, 2026.

[8] Roberto Kaban, Andi Ikmal Rachman, Indah Purnama Sari, Andi Zulherry, Mhd Basri, Nur Fadilah, Nurul Maulidiyah, David JM Sembiring, Gelard Untirtha Pratama, Ronal Ronal, Rizki Yustisia Sari, Riah Ukur Ginting, Mardhiyatna Mardhiyatna, Ahyuna Ahyuna. Pengantar Artificial Intelligence, Machine Learning, dan Big Data, Medan: Yayasan Kita Menulis, 2026.

[9] AP Irfan, Cucut Susanto, Indah Purnama Sari, Andi Zulherry, Yusron Abda'u Ansya, Idah Kusuma Dewi, Khairul Muttaqin, Rahman Pradipta, Janner Simarmata, Ahmad Ihsan, Maisaroh Maisaroh, Bayu Waseso, Roberto Kaban, Dafrid Cahyadi Arifin, Syukriy Abdullah, Adam M Tanniewa, Muharman Lubis, Inkreswari Retno Hardini, Chicha Rizka Gunawan, Roy Mubarak. Sistem Informasi Modern Untuk Era Digital, Medan: Yayasan Kita Menulis, 2026.

[10] S Wahyuddin, Indah Purnama Sari, Samuel Hasudungan Tampubolon, Andi Zulherry, Butsiarah Butsiarah, Taufik Taufik, George Morris William Tangka, Rahmad Bahri, Mhd Basri, Raemon Syaljumairi, Robbi Rahim, Muharman Lubis, Baso Ali, Sirmayanti Sirmayanti, Janner Simarmata. Cloud Computing : Referensi Lengkap, Medan: Yayasan Kita Menulis, 2025.

Journal Article from the Internet

[11] Sari, I.P., Hariani, P.P., Al-Khowarizmi, A., Ramadhani, F., Sulaiman, O.K., Satria, A, & Manurung, A.A. (2024). CLUSTERING HIV/AIDS DISEASE USING K-MEANS CLUSTERING ALGORITHM. Proceeding International Seminar on Islamic Studies 5 (1), 1668-1676

[12] Sari, I.P., Ramadhani, F., Satria, A., & Sulaiman, O.K. Leukocoria Identification: A 5-Fold Cross Validation CNN and Adaboost Hybrid Approach. 2023 6th International Seminar on Research of Information Technology and Intelligent Systems (ISRITI), 486-491

[13] Manurung, A.A., Nasution, M.D., & Sari, I.P. (2023). Implementation of Fuzzy K-Nearest Neighbor Method in Dengue Disease Classification. 2023 11th International Conference on Cyber and IT Service Management (CITSM), 1-4

[14] Sari, I.P., Ramadhani, F., Satria, A., & Apdilah, D. (2023). Implementasi Pengolahan Citra Digital dalam Pengenalan Wajah menggunakan Algoritma PCA dan Viola Jones. Hello World Jurnal Ilmu Komputer 2 (3), 146-157

[15] Zulherry, A., Basri, M., & Yusnandar, W. (2025). Implementation of Private Cloud Infrastructure as a Service Using Proxmox VE Based on the Network Development Life Cycle (NDLC) Approach. Altafani: Jurnal Pengabdian Masyarakat, 1(1).

[16] Zulherry, A., Sari, I. P., & Basri, M. (2026). Perancangan dan Implementasi Segmentasi LAN pada Infrastruktur Jaringan Skala Menengah. Hello World Jurnal Ilmu Komputer, 4(4).

[17] Zulherry, A., Riadi, I., & Umar, R. (2026). Anomaly Detection in Cloud Device-Based Information Technology Infrastructure Using Isolation Forest Algorithm. Journal Of Informatics And Telecommunication Engineering 9 (2)

[18] Zulherry, A., & Gunawan, M. (2025). Development of an Android-Based Smart Health Monitoring Device for Heartbeat Detection. Al'adzkiya International of Computer Science and Information Technology (AIoCSIT) Journal, 6(2), 75-79.

[19] Sari, I.P., Zulherry, A., Basri, M., & Hayani W. (2025). Pembelajaran Pemrograman berbasis Machine Learning sebagai Upaya Peningkatan Computational Thinking. Jurnal Penelitian, Pendidikan dan Pengajaran: JPPP 6 (3), 245-250

[20] Sari, I.P., Al-Khowarizmi, A, Sulaiman, O.K., & Apdilah, D. (2023). Implementation of Data Classification Using K-Means Algorithm in Clustering Stunting Cases. Journal of Computer Science, Information Technology and Telecommunication Engineering 4 (2), 402-412

[21] Sulaiman, O.K & Batubara, I.H. (2021). Implementation Data Mining For Level Analysis Traffic Violation By Algorithm Association Rule. Al'adzkiya International of Computer Science and Information Technology (AIoCSIT) Journal 2 (2), 128-135

[22] Sari, I.P., Batubara, I.H., & Al-Khowarizmi, A. (2021). Sensitivity Of Obtaining Errors In The Combination Of Fuzzy And Neural Networks For Conducting Student Assessment On E-Learning. International Journal of Economic, Technology and Social Sciences (Injects) 2 (1), 331-338

[23] Sari, I.P., Al-Khowarizmi, A., & Batubara, I.H. (2021). Cluster Analysis Using K-Means Algorithm and Fuzzy C-Means Clustering For Grouping Students' Abilities In Online Learning Process. Journal of Computer Science, Information Technology and Telecommunication Engineering 2 (1), 139-144

[24] Apdilah, D., & Sari, I.P. (2021). Optimization Of The Fuzzy C-Means Cluster Center For Credit Data Grouping Using Genetic Algorithms. Al'adzkiya International of Computer Science and Information Technology (AIoCSIT) Journal 2 (2), 156-163

[25] Rangkuti, A.R.S., & Sari, I.P. (2026). Penggunaan Fitur Warna Cielab Dalam Segmentasi Kematangan Buah Kakao Dengan Algoritma K-Means Clustering. Hello World Jurnal Ilmu Komputer 5 (1), 36-46

[26] Bisono, A. T., & Zulherry, A. (2025). Analisis sentimen game Genshin Impact untuk mengetahui reaksi dan harapan pemain menggunakan metode Naïve Bayes. sudo Jurnal Teknik Informatika, 4(2), 183-193.

[27] Basri, M., & Zulherry, A. (2025). Analysis of the Impact of Gambling and Online Loans in the Perspective of Informatics, Islam, and Kemuhammadiyahan. Ar-Rasyid: Jurnal Pendidikan Agama Islam, 5(1), 65-73.

[28] Asadel, A., & Zulherry, A. (2025). Detecting Zero-Width Characters Obfuscated in Phishing URLs using the XGBOOST Algorithm. Hanif Journal of Information Systems 3 (1), 43-53

[29] Zulherry, A., Sari, I.P., & Basri, M. (2025). Perancangan Aplikasi Monitoring Kehadiran Pegawai Menggunakan RFID. sudo Jurnal Teknik Informatika 4 (4), 378-384
[30] Ichsan, A., Zulherry, A., Lubis, T. A., & Shahnaz, B. A. Z. (2025). Utilization of Mobile Applications to Speed Up The Search for Android-Based Index Places. IJATCoS: Indonesian Journal of Applied Technology. Computer and Science, 2(1).
[31] Sari, I.P., Sulaiman, O.K., Apdilah, D., & Simanjuntak, P. (2025). Analisis Komparatif Support Vector Machine dan Random Forest untuk Deteksi Email Phishing. Applied Information Technology and Computer Science (AICOMS) 4 (2), 18-27
[32] Ardini, R.D., & Sari, I.P. (2026). Implementasi Algoritma Support Vector Machine (SVM) Dan Logistic Regression Dalam Menganalisis Sentimen Tanggapan Masyarakat Terhadap Tiktok Shop Di Social Media. Hello World Jurnal Ilmu Komputer 5 (1), 23-35
[33] Sari, I.P., Zulherry, A., & Basri, M. (2025). Pembelajaran Pemrograman berbasis Machine Learning sebagai Upaya Peningkatan Computational Thinking. Jurnal Penelitian, Pendidikan dan Pengajaran: JPPP 6 (3), 245-250
[34] Sitompul, M.FA., & Sari, I.P. (2026). Implementasi Algoritma Dijkstra Dalam Penentuan Rute Distribusi Beras SPHP Di Perum Bulog Kanwil Sumut. Hello World Jurnal Ilmu Komputer 4 (4), 218-225
[35] Al Wahya, B.R., & Sari, I.P. (2026). Perancangan Sistem Penjadwalan Kegiatan Belajar Mahasiswa Berbasis AI Menggunakan Agoritma Genetika Di Fakultas Ilmu Komputer Dan Teknologi Informasi. Hello World Jurnal Ilmu Komputer 4 (4), 185-195
[36] Sari, I.P., Basri, M., & Syafrayani, P.R. (2025). Implementasi Sistem Aplikasi Pengolahan Teks pada Gambar Menggunakan Modifikasi Metode LSB dan ROT13. Hello World Jurnal Ilmu Komputer 4 (2), 90-97
[37] Sari, I.P., Fadlil, A., & Sutikno, T. (2025). Islamic sound recognition using MFCC and SVM: Case study on takbir and sholawat. Journal of Computer Networks, Architecture and High Performance Computing 7 (3), 825-839
[38] Safira, T.Y., & Sari, I.P. (2025). Tingkat Akurasi dalam Analisis Perbandingan Metode Oreste dengan Preference Selection Index (PSI) Menentukan Pestisida Terbaik terhadap Hama Padi Desa Kolam Kecamatan Percut Sei Tuan. Blend Sains Jurnal Teknik 3 (3), 237-245
[39] Amada, E.P., & Zulherry, A. (2025). Klasterisasi Minat Dan Bakat Siswa Menggunakan Metode X-Means Berbasis Web: Studi Kasus SMA Negeri 1 Hamparan Perak. sudo Jurnal Teknik Informatika 4 (4), 276-283.
[40] Zulherry, A., Ramadhani, F., & Satria, A. (2024). Klasifikasi Data Tracer Study Dengan Pemanfaatan Data Mining Menggunakan Algoritma Support Vector Machine dan Neural Network. Portal Riset dan Inovasi Sistem Perangkat Lunak 2 (1), 45-54
[41] Zulherry, A. (2023). Decision making for network security with simple additive weighting method. Journal of Intelligent Decision Support System (IDSS), 6(3), 155-159.
[42] Zulherry, A., Gunawan, T. S., & Wanayumini, W. (2021). Analisis Hasil Pendukung Keputusan Mendapatkan Rumah Dinas Perusahaan Menggunakan Metode Analytical Hierarchy Process (AHP) dan Technique for Order Preference by Similarity to Ideal Solution (TOPSIS). Jurnal Media Informatika Budidarma, 5(2), 695-704.
[43] Zulherry, A., Basri, M., & Yusnandar, W. (2025). Implementation of Private Cloud Infrastructure as a Service Using Proxmox VE Based on the Network Development Life Cycle (NDLC) Approach. Altafani: Jurnal Pengabdian Masyarakat, 1(1).
[44] Dongoran, D., & Sari, I.P. (2025). Implementasi Klasifikasi Data Tracer Study Pada Universitas Muhammadiyah Sumatera Utara Dengan Pemanfaatan Data Mining Menggunakan Kombinasi Algoritma Support Vector Machine. Hello World Jurnal Ilmu Komputer 4 (1), 12-24
[45] Aqta, D.P., & Sari, I.P. (2025). Evaluasi Keefektifan Teknik Morfologi dan Histogram Citra Digital pada Minyak RON 92 di SPBU Pertamina Medan Tembung. Hello World Jurnal Ilmu Komputer 3 (4), 161-170
[46] Sari, I.P., Sulaiman, O.K., & Apdilah, D. (2025). Penerapan Sistem Pakar Diagnostik Penyakit Kelapa Sawit sebagai Solusi bagi Petani dalam Meningkatkan Produktivitas Perkebunan. CivicAction: Jurnal Pengabdian dan Inovasi Masyarakat 1 (1), 8-17