The 2nd International Conference on Multidisciplinary Engineering (2nd ICOMDEN 2024)

Proceedings of International Conference on Multidisciplinary Engineering (ICOMDEN): 00025 (2024)
https://doi.org/10.29103/icomden.v2.xxxx
Volume 2, 2024
eISSN: xxxx-xxxx
Research Original Article

Implementation Of The Adaboost Method On Linear Kernel Svm For Classifying Pip Assistance Recipients At SMP Negeri 2 Kejuruan Muda

Muhammad Fahri Al-Fikri[1], Asrianda [2], Zahratul Fitri [3]
[1,2,3] Universitas Malikussaleh
[1,2,3] muhammad.190170163@mhs.unimal.ac.id, asrianda@unimal.ac.id, zahratulfitri@unimal.ac.id
* Correspondence: Muhammad Fahri Al-Fikri

Abstract: *This study examines the application of the AdaBoost algorithm to a Linear Kernel Support Vector Machine (SVM) for determining student eligibility for the Indonesian Smart Program (PIP) at SMP N 2 Kejuruan Muda. The main objective is to improve the accuracy and fairness of the PIP aid distribution using advanced machine learning techniques. The dataset used comprises 500 student records, which include demographic, academic, and economic factors. The dataset was divided into training and testing sets, with the AdaBoost algorithm applied to enhance the SVM model's performance. The study found that the SVM model optimized with AdaBoost was able to classify 91 students as eligible for PIP aid, achieving an impressive accuracy rate of 97.85%. Only 2 students were classified as ineligible, representing 2.15% of the total sample. When compared to the standard SVM model, which also classified 91 students as eligible, the key advantage of AdaBoost lies in its ability to handle borderline data more effectively. AdaBoost improves the classification of students whose eligibility was less clear by reinforcing the importance of difficult-to-classify instances. The model's higher precision on edge cases indicates that AdaBoost offers a significant improvement over traditional SVM models in handling complex classification tasks. This research concludes that incorporating AdaBoost into SVM models provides a more robust and accurate method for determining student eligibility for government aid programs such as PIP.*

Keywords: *AdaBoost, SVM, Indonesian Smart Program, PIP aid, machine learning, student eligibility, classification.*

1. Introduction

Education serves as a crucial pillar for developing the character and intellectual capabilities of a nation's human resources. In Indonesia, the Government's Indonesia Pintar Program (PIP) aims to enhance access to quality education for economically disadvantaged students. However, a significant challenge arises in ensuring that this assistance reaches students who genuinely meet the eligibility criteria [1]. Current selection and classification processes for PIP assistance at SMP Negeri 2 Kejuruan Muda often rely on manual methods, leading to issues such as subjective bias, administrative errors, and inefficiencies that hinder accurate distribution of aid.

To address these challenges, machine learning technologies offer promising solutions through automated and objective classification algorithms. The AdaBoost method, a form of ensemble learning, has demonstrated effectiveness in enhancing prediction accuracy by combining multiple weak learners into a robust model [2]. This research integrates AdaBoost with a Linear Kernel Support Vector Machine (SVM) to leverage the strengths of both algorithms: AdaBoost's efficiency in managing complex and imbalanced data and Linear Kernel SVM's strong performance in high-dimensional feature spaces, particularly for binary classification. This integration aims to improve accuracy and reliability in determining PIP eligibility for students at SMP Negeri 2 Kejuruan Muda.

The objective of this study is to implement AdaBoost and Linear Kernel SVM in the classification process for PIP assistance eligibility at SMP Negeri 2 Kejuruan Muda. This case study is particularly relevant due to the school's unique data characteristics, which are representative for testing algorithm effectiveness. Furthermore, the collaboration between researchers and the school facilitates the collection of accurate and pertinent data. Ultimately, the research not only seeks to enhance the practical aspects of PIP assistance selection but also aims to contribute theoretically to the advancement of knowledge, particularly regarding the application of machine learning in education. By emphasizing the importance of this research, the study highlights how the implementation of AdaBoost can potentially improve the effectiveness and efficiency of the selection process while providing new insights into the use of information technology in education.

2. Literature Review

2.1. Indonesia Pintar Program (PIP)

The Indonesia Pintar Program (PIP) is an initiative by the Indonesian government aimed at providing financial assistance to students from low-income families. Its primary objective is to enhance access to quality education and reduce school dropout rates by offering direct aid for educational expenses such as tuition fees, books, uniforms, and other school necessities. PIP serves as a significant effort by the government to promote inclusive and equitable education across all segments of society [3].

2.2. Eligibility Criteria for PIP Assistance

Eligibility for PIP assistance is typically determined by several factors, including the economic status of the family, social background, and the academic performance of the student. The government establishes specific guidelines that students must meet to qualify for benefits, such as maximum family income thresholds, proof of financial need, and satisfactory academic progress reports. These criteria ensure that assistance is directed to students who genuinely require support and have the potential to improve their academic achievements through financial aid [4].

2.3. Machine Learning and Data Classification

Definition of Machine Learning

Machine learning is a branch of artificial intelligence (AI) that enables systems to learn and improve from experience without being explicitly programmed. It involves developing algorithms that can identify patterns in data and make predictions or decisions based on that data. Machine learning is applied in various domains, from product recommendations to speech recognition and natural language processing [5].

2.4. Implementation of Machine Learning in Data Classification

Classification is a machine learning task aimed at predicting the class label of a sample based on its known attributes. In the educational context, classification can be employed to determine whether a student qualifies for assistance based on characteristics such as academic performance, economic background, and other relevant factors [6].

2.5. Implementation of AdaBoost

Basic Concepts of AdaBoost

Adaptive Boosting (AdaBoost) is one of the most popular and effective ensemble learning methods. It works by combining multiple weak classifiers to create a strong classifier. Each weak classifier is constructed based on features from the provided dataset, and subsequent classifiers are built by placing more emphasis on observations that were misclassified by earlier classifiers. This iterative process improves classification accuracy[7].

3. Research Methodology

This study was conducted in Aceh Tamiang, focusing on data collection from students at SMP Negeri 2 Kejuruan Muda. The collected data includes demographic information, academic performance, family economic status, and prior PIP assistance, all aimed at classifying eligibility for PIP assistance. A structured methodology was employed, beginning with comprehensive data collection through primary sources at the school, followed by basic statistical processing to derive insights. The research process also involved system design using context diagrams, Data Flow Diagrams (DFD), and Entity-Relationship Diagrams (ERD) to guide application development.

The research methodology encompasses several critical steps aimed at enhancing the classification accuracy for PIP assistance. Initially, the study identifies the key issues surrounding PIP eligibility classification and sets specific objectives to improve accuracy using the AdaBoost algorithm. A thorough literature review on AdaBoost and machine learning classification systems is conducted, establishing a solid foundation for the research. Subsequently, quantitative methods are applied, leveraging data from students at SMP Negeri 2 Kejuruan Muda, ensuring the research is grounded in practical, real-world scenarios.

To implement the classification system, the research incorporates data cleaning and preprocessing techniques, ensuring data readiness for machine learning algorithms. The AdaBoost and Linear Kernel SVM models are developed and trained using the prepared dataset. Model evaluation is conducted using performance metrics such as accuracy, precision, recall, and F1-score to ascertain the effectiveness of the classification. The findings aim to provide insights into the factors influencing PIP eligibility classification, ultimately contributing to more objective and data-driven educational decision-making within the context of the Indonesia Pintar Program.

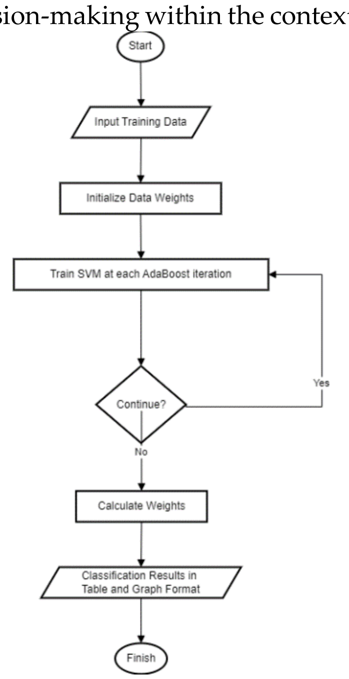

Figure 1. Research methodology

3.1. Mathematical Formulation of AdaBoost

3.1.1. Data Weight Initialization

$$D_t(i) = \frac{D_{t-1}(i) \cdot exp\left(-\alpha_t y_i h_t(x_i)\right)}{Z_t} \tag{2.1}$$

Where:

$D_t(i)$ *: weight of data at iteration* t

α_t *: learning rate at iteration* t

y_i *: class label of data point* i

$h_t(x_i)$ *: prediction of the model at iteration* t *for data point* i

Z_t*: normalization factor*

Error Rate Calculation:

$$\epsilon_t = \frac{\sum_{i=1}^{N} D_t(i) \cdot 1(h_t(x_i) \neq y_i)}{\sum_{i=1}^{N} D_t(i)} \tag{2.2}$$

Where:

1 (condition): indicator function

3.1.2. Weight Calculation for Weak Classifier

$$\alpha_t = \frac{1}{2} ln\left(\frac{1-\epsilon_t}{\epsilon_t}\right) \tag{2.3}$$

Where:

α_t *: weight for weak classifier* h_t

Weight Update for Next Iteration:

$$D_{t+1}(i) = \frac{D_t(i) \cdot exp\left(-\alpha_t y_i h_t(x_i)\right)}{Z_t} \tag{2.4}$$

3.1.3. Final Classifier Output

$$H(x) = sign\left(\sum_{t=1}^{T} \alpha_t h_t(x)\right) \tag{2.5}$$

3.2. Linear Kernel SVM

The linear kernel in Support Vector Machines (SVM) is a type of kernel function used for classifying linearly separable data. It is suitable when data can already be separated linearly. The linear kernel is one of several kernel functions available in SVM, with others including radial basis function (RBF) and polynomial kernels [8].

Mathematical Formulation for Linear Kernel SVM:

$$f(x) = sign\left(\sum_{i=1}^{n} \alpha_i y_i \langle x, x_i \rangle + b\right) \tag{2.6}$$

Where:

$f(x)$ *: prediction function*
n *: number of training data*
α_i *: weight of the support vector*
y_i *: class label of training data*
x *: feature vector to be predicted*
x_i *: feature vector from training data*
$\langle x, x_i \rangle$ *: dot product of* x *and* x_i
b: bias

Hinge Loss Function:

$$L(y, f(x)) = \max(0, 1 - y \cdot f(x)) \tag{2.7}$$

Where:
$L(y, f(x))$ *: loss function*
y *: class label*

3.3. Research Results

This study focused on the development and evaluation of Support Vector Machine (SVM) models, both in their standard form and optimized with the AdaBoost algorithm, to classify the eligibility of students for the Indonesia Pintar Program (PIP) assistance. Utilizing a dataset comprising over 500 training samples, the research aimed to optimize the classification performance by adjusting weights and enhancing the model with AdaBoost techniques. The results of the model were visualized using Principal Component Analysis (PCA), allowing for an insightful examination of the data structure and the influence of various features.

1. SVM Model Formation Results

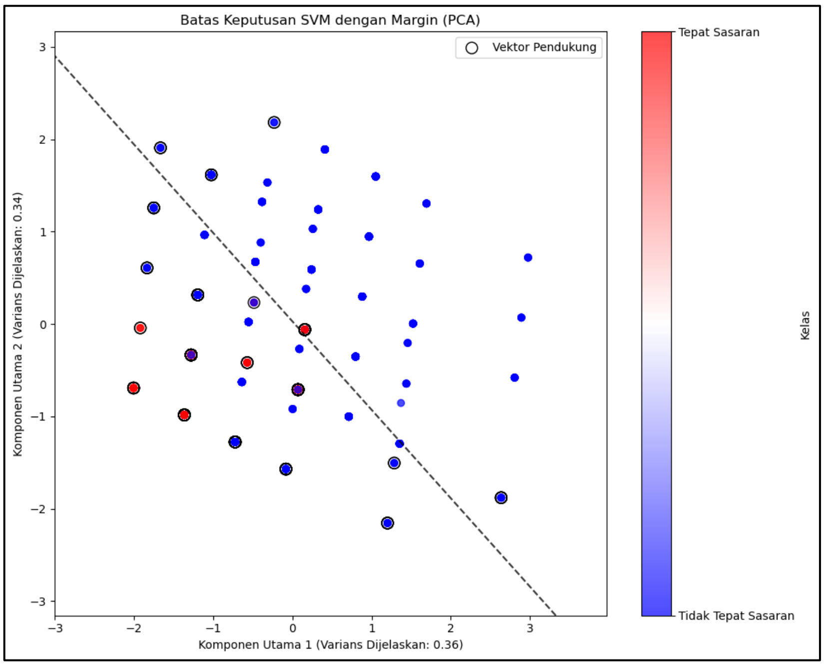

Figure 2. Model formation results

The PCA plot displayed in illustrates the results of the SVM model. The analysis revealed two primary components that significantly influenced the classification outcome:

Principal Component 1 (PC1): This component is primarily influenced by two features:

- **Income (bobot_penghasilan)**: This feature has a weight of -0.6853, indicating a strong negative correlation with this component. Higher income values correspond to lower positions along this axis, suggesting that students with lower incomes are more likely to be classified as eligible for PIP assistance.
- **Previous Assistance Received (bobot_pernah_menerima)**: With a weight of -0.7178, this feature also exhibits a strong negative influence on PC1. The proximity of data points along this axis indicates that students who have previously received assistance are clustered together, emphasizing their continued need for support.

Principal Component 2 (PC2): This component is dominated by the feature "Dependents (bobot_tanggungan)" which carries a weight of -0.9400. This suggests that the number of dependents significantly impacts the vertical data distribution. Students with a higher number of dependents are positioned further from the center along this axis, highlighting their greater need for assistance. Additionally, the income feature contributes positively to PC2 with a weight of 0.3124, although its effect is less pronounced compared to dependents.

The PCA visualization indicates a clear separation of classes based on income and previous assistance, revealing how these factors contribute to the eligibility classification for PIP assistance.

2. SVM Model with AdaBoost Optimization

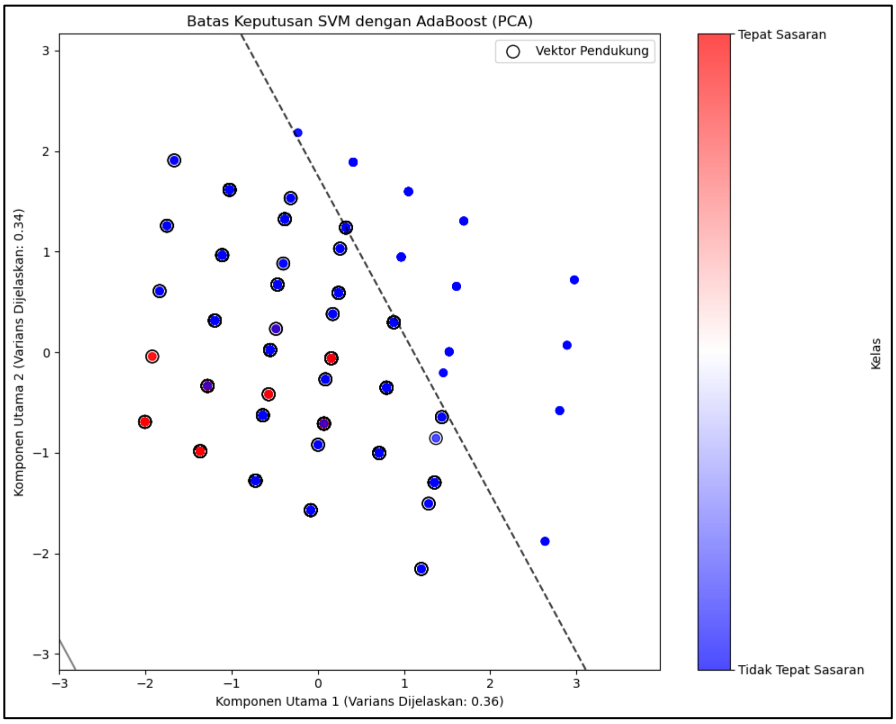

Figure 3. SVM model with adabost optimazation

In Figure the optimization of the SVM model through the implementation of AdaBoost is presented. The key findings from this visualization are as follows:

- **Decision Boundary (Hyperplane)**: The dashed line represents the optimized hyperplane that separates the two classes (eligible and ineligible for PIP assistance). AdaBoost enhances the positioning of this boundary by focusing on data points that are difficult to classify correctly. This is achieved by increasing the weights on misclassified examples, enabling the model to improve its accuracy over iterations.

- **Impact of AdaBoost**: The addition of AdaBoost emphasizes the importance of harder-to-classify data points. Compared to the standard SVM model, AdaBoost improves classification outcomes by adjusting weights for examples near the margin, resulting in a more robust decision boundary. This leads to a reduction in misclassification rates, particularly for points that were previously misclassified by the SVM without optimization.
- **Support Vectors**: The circled points on the plot represent the support vectors, which are crucial in determining the position and orientation of the hyperplane. With AdaBoost optimization, the relevance of these support vectors increases, as they help the model learn from challenging cases more effectively. The visualization clearly shows that these points are positioned close to the decision boundary, highlighting their significance in the classification process.
- **Data Distribution**: The distribution of data along PC1 and PC2 appears more distinct with the integration of AdaBoost. Points representing the two classes, 'Targeted' and 'Not Targeted', show improved separation, even in the presence of outliers. AdaBoost effectively addresses classification errors related to outliers, ensuring a clearer delineation between eligible and ineligible candidates.

3. Comparison Between SVM Models

Table 1. Comparison of SVM and SVM with AdaBoost Models

Criteria	SVM	SVM with AdaBoost
Decision Boundary (Hyperplane)	Based on the largest margin, but some data points are difficult to separate clearly.	Improved decision boundary with emphasis on misclassified data points near the margin.
Support Vectors	Data points near the margin determine hyperplane position, with some potentially suboptimal selections.	Selected support vectors are more relevant due to AdaBoost's focus on difficult cases.
Data Distribution	Data generally well-distributed; however, some outliers affect predictions.	Data shows clearer separation due to AdaBoost's correction of errors related to outliers.
Classification Accuracy	Standard accuracy; some points near the margin may be misclassified.	Improved accuracy, especially on hard-to-classify points near the margin.
Confidence in Predictions	Predictions are adequate for most data, but confidence is lower near the margin.	Increased confidence, particularly for data previously close to the margin or misclassified.
Handling Outliers	Standard SVM struggles with outliers close to the margin.	Outliers are better managed as AdaBoost corrects recurring misclassification errors.
Computational Time	Faster due to the lack of additional boosting iterations.	Longer computation time due to repeated iterations for accuracy improvement.

Summary of Classification Results

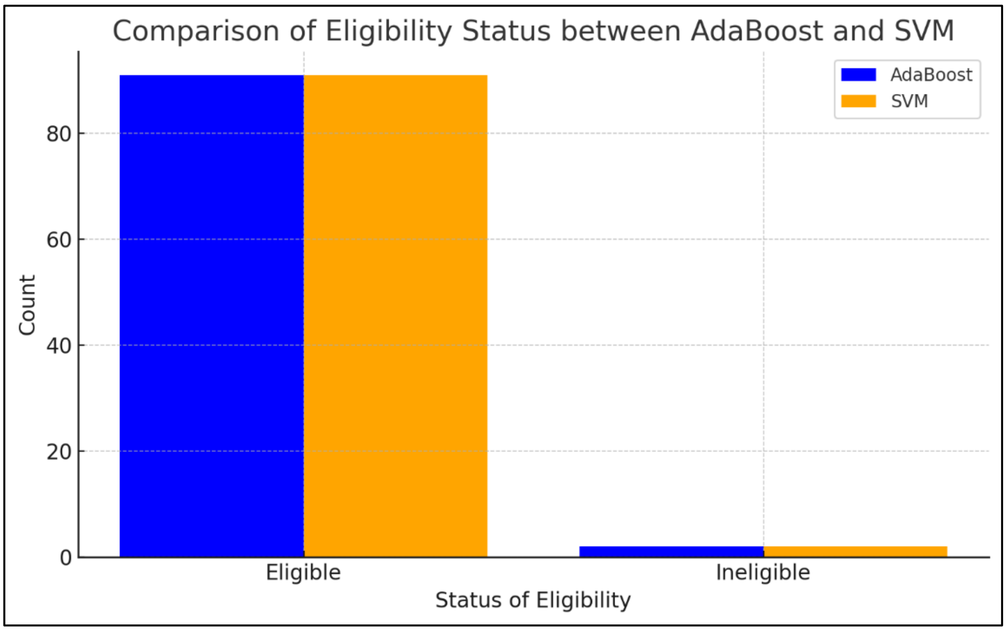

Figure 4. Summary of clasification results

Table 2. The final classification results

Description	Count	Percentage (%)
Eligible (SVM with AdaBoost)	91	97.85%
Ineligible (SVM with AdaBoost)	2	2.15%
Eligible (SVM)	91	97.85%
Ineligible (SVM)	2	2.15%

In the study, the implementation of the SVM model enhanced with AdaBoost produced significant results regarding student eligibility for PIP assistance. Specifically, a total of 91 students were deemed eligible by the SVM model using AdaBoost, which translates to an impressive 97.85% of the analyzed dataset classified as eligible for assistance. In contrast, only 2 students were identified as ineligible by this model, representing 2.15% of the total data.

When examining the results from the standard SVM model, the findings were notably consistent. The same number of students—91—was also classified as eligible, maintaining the percentage at 97.85%. Similarly, the standard SVM model identified 2 students as ineligible, which again accounted for 2.15% of the analyzed dataset.

These results indicate a high level of agreement between the two models in classifying student eligibility for PIP assistance, showcasing the effectiveness of both the SVM and SVM with AdaBoost in accurately identifying those who qualify for support.

4. Conclusions

The implementation of Support Vector Machine (SVM) and the optimized version of SVM with AdaBoost in this study effectively classified recipients of the Indonesia Pintar Program (PIP) assistance. The SVM was used as the baseline model with a linear kernel to separate data based on the largest possible margin, utilizing a training dataset of 638 entries and a test dataset of 93 entries. The features used for training included income, economic status, number of dependents, and assistance status. Following this, AdaBoost was applied to enhance the model's performance by increasing the weights of misclassified data points during previous iterations. This approach allowed the model to learn more complex patterns and make more accurate classifications, particularly focusing on difficult-to-classify data points, thereby reducing classification errors.

The results demonstrated that both the pure SVM model and the SVM optimized with AdaBoost yielded identical outcomes regarding classification. Out of 93 test data points, 91 were classified as "Eligible," equating to 97.85%, while only 2 were deemed "Ineligible" (2.15%). This indicated an excellent performance by both models in detecting eligible and ineligible students, achieving an accuracy rate of 97.85%. While the addition of AdaBoost did not significantly alter the overall accuracy, it provided notable technical advantages. Specifically, AdaBoost improved the handling of outliers and data points close to the decision margin by enhancing their weights and refining the support vectors. This optimization led to better class separation and increased confidence in predictions for challenging data points. However, it is important to note that the implementation of AdaBoost increased computational time due to the additional iterations required. Ultimately, even though there was no significant difference in overall accuracy, SVM with AdaBoost demonstrated improvements in predictive accuracy for more complex data, better handling of outliers, and a more stable overall model.

References

[1] M. Fiqih, U. Dayat, and K. Febriantin, “Implementasi kebijakan program indonesia pintar (pip) pada jenjang sekolah dasar,” Kinerja, vol. 18, no. 4, pp. 539–546, 2022, doi: 10.30872/jkin.v18i4.10097.

[2] U. Indahyanti, N. L. Azizah, and H. Setiawan, “Pendekatan Ensemble Learning Untuk Meningkatkan Akurasi Prediksi Kinerja Akademik Mahasiswa,” J. Sains dan Inform., vol. 8, no. 2, pp. 160–169, 2022, doi: 10.34128/jsi.v8i2.459.

[3] M. Fiqih, U. Dayat, and K. Febriantin, “Implementasi kebijakan program indonesia pintar (pip) pada jenjang sekolah dasar,” Kinerja, vol. 18, no. 4. pp. 539–546, 2022. doi: 10.30872/jkin.v18i4.10097.

[4] M. S. Rahmawati, T. H. Iskandar Alam, and S. Sulfiana, “Sistem Pendukung Keputusan Seleksi Penerima Beasiswa Program Indonesia Pintar di SMK Negeri 1 Sorong Dengan Menggunakan Metode Analytical Hierarchy Process Berbasis Dekstop,” Insect (Informatics Secur. J. Tek. Inform., vol. 5, no. 1, p. 31, 2019, doi: 10.33506/insect.v5i1.1281.

[5] Mahesh Batta, “Machine Learning Algorithms - A Review,” Int. J. Sci. Res., no. October, 2020, doi: 10.21275/ART20203995.

[6] S. Saroji, E. Winata, P. P. W. Hidayat, S. Prakoso, and F. Herdiansyah, “The Implementation of Machine Learning in Lithofacies Classification using Multi Well Logs Data,” Aceh Int. J. Sci. Technol., vol. 10, no. 1, pp. 9–17, 2021, doi: 10.13170/aijst.10.1.18749.

[7] D. Gao et al., “An Intelligent Control Method for Servo Motor Based on Reinforcement Learning,” Algorithms, vol. 17, no. 1, pp. 1–15, 2024, doi: 10.3390/a17010014.

[8] A. Z. Praghakusma and N. Charibaldi, “Komparasi Fungsi Kernel Metode Support Vector Machine untuk Analisis Sentimen Instagram dan Twitter (Studi Kasus : Komisi Pemberantasan Korupsi),” JSTIE (Jurnal Sarj. Tek. Inform., vol. 9, no. 2, p. 88, 2021, doi: 10.12928/jstie.v9i2.20181.