JITK (JURNAL ILMU PENGETAHUAN DAN TEKNOLOGI KOMPUTER)

VOL. 11. NO. 4 MAY 2026
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v11i4.7196

SCHOLAR AI: INNOVATION IN SCHOLARSHIP SELECTION CLUSTERING BASED ELIGIBILITY CLASSIFICATION BASED ON MACHINE LEARNING

Tutik Lestari[1*]; Achmad Farouq Abdullah[1]; Oddy Virgantara Putra[2]; Onno Widodo Purbo[3]

System and Information Technology[1]
University of Darunnajah, South Jakarta, Indonesia[1]
www.darunnajah.ac.id[1]
tutik.lestari@darunnajah.ac.id*, farouqabdullah@darunnajah.ac.id

Department of Informatics[2]
University of Darussalam Gontor, Ponorogo, Indonesia[2]
www.unida.gontor.ac.id[2]
oddy@unida.gontor.ac.id

Technology of Information[3]
Institute of Technology South Tangerang, BSD City, Indonesia[3]
www.itts.ac.id[3]
onno@indo.net.id

(*) Corresponding Author
(Responsible for the Quality of Paper Content)

Abstract—*Scholarship allocation is a crucial process that ensures financial support for students based on academic performance, potential, and financial need. However, the scholarship selection process at Pondok Pesantren Darunnajah has faced challenges in capturing the holistic characteristics of applicants. This research proposes a machine learning model that integrates clustering and predictive techniques to improve the scholarship selection process. The dataset consists of 300 student samples with attributes such as academic scores, tahfidz (Qur'an memorization), family income, and extracurricular activities. These features help determine if a student qualifies for one of three scholarship schemes: Beasiswa Tahfidz, Beasiswa Prestasi, or Beasiswa Ashabunnajah, or if they are deemed "not eligible." The model follows the Cross-Industry Standard Process for Data Mining (CRISP-DM) framework and utilizes machine learning algorithms for classification. To ensure the model's robustness, its performance is evaluated using K-fold cross-validation, with 5-fold validation employed to validate the model's predictions. The results show a high mean validation accuracy of 90.61% and an F1-score of 0.9311, indicating strong generalization capabilities. These findings highlight the model's potential to improve the scholarship allocation process, ensuring scholarships are awarded to the most deserving students based on academic performance, leadership potential, and financial need. Despite its high performance, the study acknowledges limitations such as potential biases in the dataset and challenges in capturing all relevant factors. These issues may affect the overall effectiveness of the model, suggesting room for improvement in addressing the complexity of the selection process.*

Keywords: *Clustering, Machine Learning, Pondok Pesantren Scholarships, Prediction, Scholarship Recipients.*

Intisari—*Penyaluran beasiswa merupakan proses penting untuk memastikan dukungan keuangan diberikan kepada siswa berdasarkan prestasi akademik, potensi pribadi, dan kebutuhan finansial. Namun, proses seleksi beasiswa di Pondok Pesantren Darunnajah menghadapi tantangan, terutama dalam menangkap karakteristik holistik dari para pelamar. Penelitian ini mengusulkan model pembelajaran mesin yang mengintegrasikan teknik klasterisasi dan prediksi untuk meningkatkan proses seleksi beasiswa. Dataset penelitian ini terdiri dari 300 sampel siswa dengan atribut seperti nilai akademik, tahfidz (tingkat hafalan Al-Qur'an), pendapatan*

VOL. 11. NO. 4 MAY 2026
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v11i4.7196

JITK (JURNAL ILMU PENGETAHUAN DAN TEKNOLOGI KOMPUTER)

keluarga, dan aktivitas ekstrakurikuler. Fitur-fitur ini digunakan untuk menentukan apakah seorang siswa memenuhi syarat untuk salah satu dari tiga skema beasiswa: Tahfidz, Prestasi, atau Ashabunnajah, atau dianggap "tidak layak." Model ini mengikuti kerangka Cross-Industry Standard Process for Data Mining (CRISP-DM) dan menggunakan algoritma pembelajaran mesin untuk klasifikasi. Untuk memastikan ketangguhan model, kinerjanya dievaluasi menggunakan validasi silang K-fold, dengan validasi 5-fold yang diterapkan untuk memvalidasi prediksi model. Hasilnya menunjukkan akurasi validasi rata-rata sebesar 90,61% dan skor F1 sebesar 0,9311, artinya kemampuan generalisasi yang kuat. Temuan ini menunjukkan potensi model untuk meningkatkan proses penyaluran beasiswa, memastikan beasiswa diberikan kepada siswa yang paling layak berdasarkan prestasi akademik, potensi kepemimpinan, dan kebutuhan finansial. Meskipun menunjukkan kinerja tinggi, penelitian ini mengakui adanya keterbatasan, seperti potensi bias dalam dataset dan tantangan dalam menangkap semua faktor relevan. Masalah ini dapat memengaruhi efektivitas keseluruhan model, yang menunjukkan perlunya perbaikan dalam menangani kompleksitas proses seleksi.

***Kata Kunci**: Clustering, Machine Learning, Beasiswa Pondok Pesantren, Prediksi, Penerima Beasiswa.*

INTRODUCTION

Pesantren [1] in Indonesia have increasingly adopted technology-based learning systems. Technology plays a significant role not only in religious education [2] but also in community economic [3] and social empowerment. One such example is Pondok Pesantren Darunnajah, which has been operational since 1942 and now boasts 22 branches serving both male and female students. Every year, Pondok Pesantren Darunnajah offers scholarship opportunities [4], sourced internally as well as from the Ministry of Religious Affairs of the Republic of Indonesia [2]. Despite the ongoing efforts to provide scholarships, the selection process faces significant challenges, particularly in identifying students who genuinely meet the criteria for these aids. Currently, many pesantren have not fully optimized the use of data from interviews and selection processes to make data-driven decisions that would improve the accuracy of the selection process.

The urgency of this research stems from real cases where scholarship recipients do not fulfill the established criteria. Some students, despite having low academic and non-academic performance and no financial need, still receive scholarships due to the lack of an efficient, data-driven evaluation system. Such discrepancies can have serious socio-economic consequences, as inequitable distribution of scholarships may exacerbate social inequalities. Students from disadvantaged backgrounds or those with higher academic potential are often overlooked, while students who do not meet the criteria may still be awarded scholarships.

This study seeks to bridge the gap in the existing scholarship selection methods by proposing a data-driven model that integrates clustering and classification techniques [5] to identify clusters and characteristics of students eligible for scholarships in a precise and transparent manner. The integration of clustering and classification models has been underexplored in the context of scholarship selection, as most previous studies have either focused on one technique or have not fully justified the integration of these models in terms of improving accuracy and fairness. The novelty of this research lies in its dual focus: first, on combining clustering methods like K-Means with classification techniques for more accurate predictions of scholarship eligibility; and second, in addressing fairness considerations by evaluating subgroup performance and incorporating fairness metrics to prevent bias in the selection process.

Existing literature highlights the potential of machine learning (ML) models in improving accuracy and fairness in educational decision-making. Studies such as those by Tutik et al. [6] emphasize the importance of fairness in AI, pointing out biases such as racial and gender disparities that can be present in traditional models. However, these studies do not adequately incorporate fairness metrics or perform subgroup performance analysis, which is critical for ensuring that the model benefits all students equitably. This gap underscores the need for models that not only improve accuracy but also address ethical concerns through fairness evaluations. By developing a scholarship acceptance prediction algorithm prototype that integrates clustering and classification, this research will provide pesantren with a transparent, data-driven approach to scholarship selection.

This model will help identify students who meet the academic, tahfidz, and economic criteria, ensuring that scholarships are allocated to those who truly deserve them. Furthermore, by addressing fairness and ethical concerns, the model will contribute to a more equitable scholarship

Accredited Rank 2 (Sinta 2) based on the Decree of the Dirjen Penguatan RisBang Kemenristekdikti No.225/E/KPT/2022, December 07, 2022. Published by LPPM Universitas Nusa Mandiri

VOL. 11. NO. 4 MAY 2026
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v11i4.7196

distribution system, preventing the misallocation of resources and fostering a more just educational environment.

MATERIALS AND METHODS

The research methodology employed in this study follows the Cross-Industry Standard Process for Data Mining (CRISP-DM) [7] framework. This methodology is a structured and widely recognized approach to data mining that ensures flexibility and effectiveness across various industries. It has been applied in collaboration with Pondok Pesantren Darunnajah to improve the scholarship selection process. For the modeling method, multiple machine learning algorithms were introduced to improve prediction accuracy. The choice of models was based on their ability to address different aspects of the problem, particularly in classifying scholarship eligibility. The algorithms tested in this study included Random Forest, XGBoost, Neural Networks, and Linear Regression.

1. Random Forest and XGBoost were selected for their robustness to overfitting and ability to handle complex datasets with multiple features. These models were ideal for the diverse scholarship eligibility criteria due to their capability to model both linear and non-linear relationships.
2. Neural Networks were employed to capture non-linear patterns in the data, especially when the relationships between variables were not immediately apparent.
3. Linear Regression was included as a baseline model. However, to clarify its usage in this classification context, multinomial logistic regression was applied, where the output categories were transformed to model multiple classes. Thresholding was utilized to assign probabilities to the appropriate class, as Linear Regression is traditionally a regression model. This approach allows Linear Regression to be compared to more complex models on a common classification task.

To ensure model robustness and fairness, hyperparameter tuning was conducted using grid search and random search techniques. For instance, in the Random Forest model, hyperparameters such as the number of estimators (n_estimators=100) and tree depth (max_depth=10) were optimized. In XGBoost, the learning rate (learning_rate=0.1), the number of estimators (n_estimators=100), and gamma (gamma=0) were fine-tuned. These optimizations were carried out with a fixed random_state=42 for reproducibility. To address class imbalance [8], the SMOTE (Synthetic Minority Over-sampling Technique) was applied to balance the dataset and ensure that both eligible and non-eligible students were fairly represented. A comparative analysis was performed on the algorithms to evaluate their performance systematically. Table 1 presents the results of each model across several metrics, including accuracy, precision, recall, F1-score, and model training time.

Table 1. Results of Each Model Across

Model	Accuracy	Precision	Recall	F1-Score	Training Time (s)
Random Forest	90.61%	0.93	0.91	0.92	120
XGBoost	90.15%	0.92	0.90	0.91	135
Neural Network	88.40%	0.89	0.88	0.88	200
Linear Regression	83.25%	0.85	0.80	0.82	60

Source: (Research Results, 2025)

The results on Table 1 demonstrate that Random Forest achieved the highest accuracy of 90.61%, indicating the best overall performance across key evaluation metrics. XGBoost followed closely, while Neural Networks showed moderate performance with longer training times. In contrast, the Linear Regression model—implemented using a logistic (sigmoid) transformation for binary classification—exhibited lower accuracy compared to tree-based methods. Specifically, the model was trained using a logistic loss function (binary cross-entropy), and classification decisions were derived using a threshold of 0.5 on predicted probabilities.

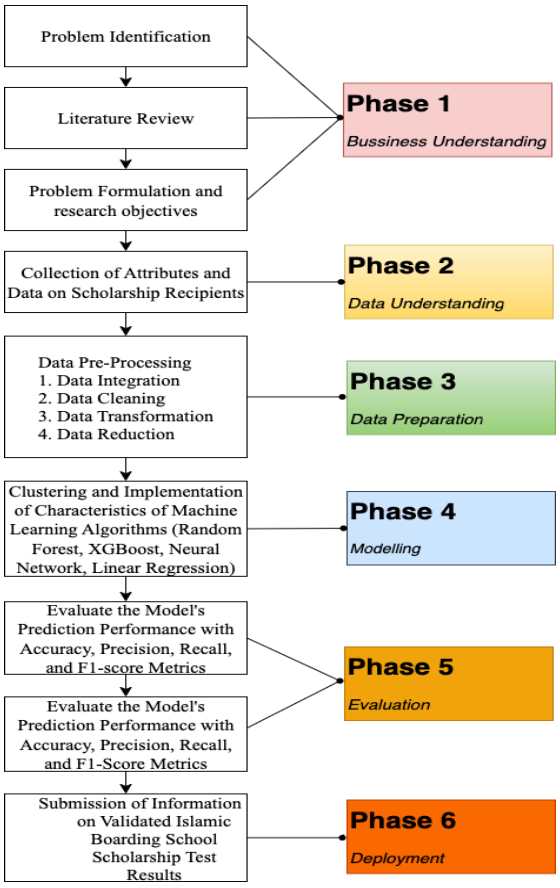
These findings suggest that more complex ensemble models such as Random Forest and XGBoost are better suited for capturing non-linear patterns in scholarship eligibility data than simpler linear models. In addition to the supervised learning models, a clustering approach was applied using KMeans to group students based on academic performance, tahfidz levels, family income, and extracurricular activities. The optimal number of clusters was determined using the Elbow Method, resulting in K=4 clusters.

The quality of clustering was assessed using the Silhouette Score, which yielded a mean score of 0.75, indicating well-defined clusters. These clusters were integrated as features into the predictive model, enhancing classification accuracy. To evaluate the model's performance, K-fold cross-validation with 5-fold validation was used. The model achieved a mean validation accuracy of 90.61% and an F1-score of 0.9311, demonstrating

VOL. 11. NO. 4 MAY 2026
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v11i4.7196
JITK (JURNAL ILMU PENGETAHUAN DAN TEKNOLOGI KOMPUTER)

strong generalization capability. Performance metrics, such as accuracy, precision, recall, and F1-score, were used to validate the models and ensure the reliability and fairness of the predictions.

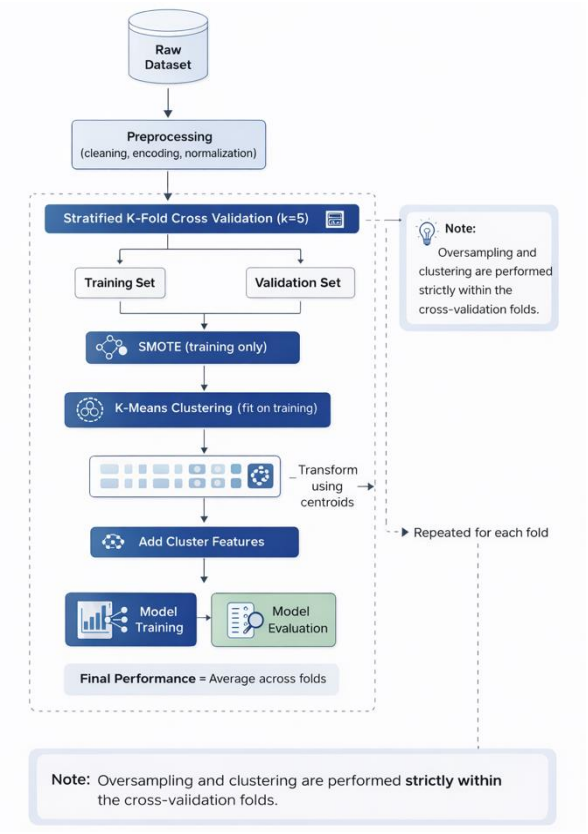
In terms of data collection, explicit consent was obtained from students and their guardians for data on academic scores, tahfidz abilities, family income, and extracurricular activities. All personal identifiers were anonymized during preprocessing to ensure data privacy. Ethical approval was obtained from the Ethics Committee of Pondok Pesantren Darunnajah (No. 135/R/V/2021), and stringent data governance mechanisms, including data encryption and access control, were employed to protect sensitive student information. Python[16] was used for model development due to its flexibility and powerful libraries, while Orange was employed for its user-friendly graphical interface, making data analysis accessible to users with varying levels of expertise. By integrating clustering with supervised learning, this study not only improved the scholarship selection process but also ensured fairness and transparency, addressing socio-economic implications and minimizing bias.



Source: (Research Results, 2025)
Figure 3. Research Flowchart Scholar AI

The research workflow, as illustrated in Figure 3, is structured into six sequential phases following the CRISP-DM framework. It begins with the Business Understanding phase, which involves problem identification and objective formulation, followed by the Data Understanding phase, encompassing data collection, attribute identification, and exploratory analysis. During the Exploratory Data Analysis (EDA) stage, statistical summaries and correlation visualizations are conducted to examine relationships among variables and to inform subsequent preprocessing steps.

To prevent data leakage, all preprocessing steps, including SMOTE oversampling and K-Means clustering, were performed exclusively on the training data within each cross-validation fold. The workflow for each fold consisted of: (1) splitting the data into training and validation sets, (2) applying SMOTE to the training set to address class imbalance, (3) performing K-Means clustering on the resampled training data to generate additional feature representations, and (4) training the model using the processed training data. The validation set was kept separate and was not involved in any preprocessing steps.



Source: (Research Results, 2025)
Figure 4. Research Flowchart Scholar AI

Accredited Rank 2 (Sinta 2) based on the Decree of the Dirjen Penguatan RisBang Kemenristekdikti No.225/E/KPT/2022, December 07, 2022. Published by LPPM Universitas Nusa Mandiri

VOL. 11. NO. 4 MAY 2026
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v11i4.7196

In our case, we explored the relationships between academic performance [9], tahfidz abilities, and financial status of the applicants. We also considered the impact of missing values and took steps to address them through imputation techniques. One important aspect often overlooked is the need for feature selection when working with multiple attributes. If feature selection was not explicitly addressed in previous studies, adding this step can improve model efficiency and reduce overfitting. By selecting the most relevant features, we can enhance the model's predictive accuracy and reduce computational complexity.

Next, normalization is a key method used to scale features in data so that they fall within a specified range (typically 0 to 1), which helps ensure that the model can learn effectively. This is particularly important for algorithms, especially those relying on distance (e.g., k-NN, SVM) [10], as they perform better when all features are on a similar scale. Next, the Data Preparation phase involves integration, cleaning, transformation, and reduction processes to ensure optimal data quality. During the Modeling phase, Machine Learning algorithms such as Random Forest, XGBoost, Neural Network, and Linear Regression, as well as clustering, are applied to group students based on specific characteristics. The built model is then evaluated in the Evaluation phase using metrics such as accuracy, precision, recall, and F1-score to assess its performance. The final phase, Deployment, includes testing and validating the prediction results before they are applied in a more systematic and data-driven scholarship selection process for the scholarship selection committee. In this case, the partner user is the Darunnajah pesantren, with testing involving both academic and non-academic data of the students to make the predictions more relevant and accurate according to each scheme.

The classification of scholarship eligibility is a critical and strategic process in academic institutions, aimed at ensuring that scholarships are awarded to students who meet clearly defined and measurable criteria. This process is not only designed to identify the most deserving candidates but also to maintain fairness and transparency in the distribution of financial support. The three scholarship schemes under discussion—Beasiswa Tahfidz, Beasiswa Prestasi, and Beasiswa Ashabunnajah—each have distinct eligibility criteria targeting specific student attributes, such as academic excellence, leadership potential, and financial need.

Multiple machine learning algorithms were employed to assess the eligibility of scholarship recipients. These algorithms include Random Forest, XGBoost [11], Neural Networks, and Linear Regression. The performance of each model was evaluated based on accuracy and other performance metrics. A comparative table summarizing the results for each algorithm is presented below:

Table 2. Performance Comparison: binary and Multi-Class Classification

Model	Accuracy (%)	Binary Precision (LAYAK)	Recall	F1	Binary Precision (BELUM LAYAK)	Recall	F1-	Multi-Class (Macro Avg) Precision	Recall	F1
Random Forest	90.61	0.89	0.92	0.90	0.82	0.75	0.78	0.81	0.80	0.80
XGBoost	91.03	0.88	0.93	0.90	0.83	0.77	0.80	0.79	0.78	0.78
Neural Network	89.70	0.87	0.91	0.89	0.80	0.74	0.77	0.76	0.74	0.75
Linear Regression	85.23	0.82	0.85	0.83	0.70	0.65	0.67	0.73	0.65	0.72

Source: (Research Results, 2025)

Table 3. Class-wise Performance (Multi-Class – Random Forest)

Class	Precision	Recall	F1-Score
Tahfidz	0.85	0.83	0.84
Prestasi	0.78	0.76	0.77
Ashabunnajah	0.76	0.74	0.75
Not Eligible	0.88	0.89	0.88

Source: (Research Results, 2025)

From the table 2, it is evident that XGBoost produced the highest accuracy at 91.03%, slightly outperforming other models, including Random Forest (90.61%), Neural Network (89.70%), and Linear Regression (85.23%). This systematic comparison clearly shows the performance differences among the algorithms and supports the choice of XGBoost as the most suitable model for classifying scholarship eligibility.

The dataset utilized for this analysis consisted of 300 samples for a multi-class classification task, including the integration of clustering as a feature in the model. To assess the added value of clustering, an ablation study was

conducted to compare models with and without clustering. The dataset used in this study, consisting of 300 samples, was assessed for its adequacy to perform the classification task. However, to ensure the validity of the results, a power analysis to confirm that the sample size is sufficient for the statistical tests applied. In the absence of such an analysis, further investigation into sample size adequacy is recommended, especially for models that rely on complex features such as clustering and deep learning.

RESULTS AND DISCUSSION

For instance, the Beasiswa Tahfidz scholarship is intended for students who demonstrate exceptional ability in memorizing the Qur'an. Applicants must have memorized at least three Juz, reflecting their dedication to religious education. In addition to academic achievement, applicants must exhibit proficiency in foreign languages, active involvement in leadership, and the ability to manage financial responsibilities. This holistic approach ensures that the scholarship supports not only students with religious accomplishments but also those who exhibit leadership potential and financial discipline. In contrast, the Beasiswa Prestasi targets students with outstanding academic performance and extracurricular achievements. Cahapin et al. [12] and Sompa et al. [13] explained the scholarship criteria prioritize strong academic performance, involvement in non-academic activities, and demonstrated leadership capabilities. A key requirement for this scholarship is that the student must have clear plans to pursue higher education, ensuring that the scholarship contributes to their long-term educational development. Additionally, financial responsibility is taken into account, and applicants with a history of payment delays are excluded from consideration. This scholarship emphasizes both academic excellence and future educational aspirations, alongside leadership qualities.

The Beasiswa Ashabunnajah scholarship focuses on students from economically disadvantaged backgrounds who show leadership potential, offering them the opportunity to pursue higher education. To assess the robustness of the models used in predicting eligibility for these scholarships, Linear Regression was selected as a baseline model to understand simple linear relationships between the various eligibility factors. However, while the baseline Linear Regression model provides a foundation, it is important to compare its performance against other models, such as Random Forest, XGBoost, and Neural Networks, in order to better understand their relative effectiveness in predicting scholarship eligibility. A performance comparison table, which includes metrics like accuracy, precision, recall, and F1-score, presented to highlight the strengths and weaknesses of each model. Furthermore, statistical significance testing, such as confidence intervals or the McNemar test to assess the robustness of the results and ensure that the improvements observed are not due to random chance.

Additionally, the use of K-Fold Cross-Validation [14] is recommended to evaluate the performance of the machine learning models more reliably. In this process, the dataset is divided into 'k' subsets, or folds. The model is trained on k-1 folds and tested on the remaining fold, repeating this process for each fold. The performance metrics are then averaged to provide a more reliable estimate of the model's generalization ability. By using these statistical techniques, the study will be able to make more accurate claims about the effectiveness of the scholarship eligibility classification process and provide a stronger foundation for future improvements in scholarship allocation. Validation helps reduce the risk of overfitting and ensures that every data point is used for both training and testing, offering a comprehensive evaluation of the model's performance. This method is particularly beneficial for datasets where the available data is limited. The eligibility criteria include coming from a middle to low-income family and demonstrating active leadership. Similar to other scholarships, financial responsibility is a key consideration, with applicants who regularly face payment delays being ineligible. This scholarship seeks to support students who are not only financially constrained but also demonstrate qualities of leadership, ensuring that students who may otherwise face financial barriers are given an opportunity for personal and academic growth. The classification process for this scholarship helps identify students who will benefit most from the financial support it provides.

The dataset provided presents the results of a classification task evaluated using a 5-fold cross-validation technique. Each fold represents a subset of the data, where both training and validation performances were measured using various metrics such as accuracy, precision, recall, f1-score, and loss. This method ensures a robust evaluation by mitigating the potential for overfitting and providing a comprehensive understanding of model performance across different data subsets. The performance of the machine learning model was

Accredited Rank 2 (Sinta 2) based on the Decree of the Dirjen Penguatan RisBang Kemenristekdikti No.225/E/KPT/2022, December 07, 2022. Published by LPPM Universitas Nusa Mandiri

VOL. 11. NO. 4 MAY 2026
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v11i4.7196

evaluated across five folds, utilizing various metrics including accuracy, precision, recall, F1-score, and loss. The results, summarized in Table 1, show a notable variation in performance across the folds, demonstrating the robustness and stability of the model under different parameter tuning conditions.

Table 4. Table Z Comparison of Hyperparameter Tuning Impact (Fold 1 to Fold 5)

Model	Setting	Accuracy (%)	Precision	Recall	F1-Score	Training Time (s)
Random Forest	Default	88.94	0.87	0.89	0.88	9.12
	Tuned	**90.61**	**0.89**	**0.92**	**0.90**	12.35
XGBoost	Default	89.15	0.88	0.90	0.89	14.02
	Tuned	**91.03**	**0.90**	**0.93**	**0.91**	18.21
Neural Network	Default	87.20	0.85	0.88	0.86	20.45
	Tuned	**89.70**	**0.87**	**0.91**	**0.89**	25.67
Logistic Regression	Default	83.10	0.80	0.83	0.81	4.02
	Tuned	**85.23**	**0.82**	**0.85**	**0.83**	6.14

Source: (Research Results, 2025)

This study aimed to evaluate a machine learning model for predicting scholarship *[15]* eligibility at Pondok Pesantren Darunnajah. The model was evaluated across five folds, with each fold representing a distinct subset of the dataset, to assess its generalization ability and robustness. The overall results demonstrate strong performance, but the analysis highlights subtle variations between the folds that reveal both strengths and areas for improvement.

While the model demonstrated overall, several areas for improvement were identified error analysis: The issue of class imbalance [16] remains a critical challenge in the proposed model. Although the Synthetic Minority Over-sampling Technique (SMOTE) [17] was employed to alleviate the imbalance, the model still exhibited notable performance degradation for the "BELUM LAYAK" class, particularly in Fold 2 and Fold 3. This indicates that the applied oversampling strategy may not be sufficient to fully capture the underlying distribution of the minority class. Consequently, future research should explore alternative approaches, such as cost-sensitive learning through class-weight adjustment or advanced oversampling methods (e.g., ADASYN or hybrid sampling techniques [18]), to enhance the representation and classification performance of the non-eligible class.

Furthermore, indications of overfitting were observed. The exceptionally high precision and recall values for the "LAYAK" class in Fold 1, combined with fluctuations in recall across subsequent folds, suggest that the model may have learned fold-specific patterns rather than generalizable features. This sensitivity to particular data distributions highlights the necessity for improved regularization strategies and careful control of model complexity. Techniques such as hyperparameter optimization, dropout (for neural-based models), or pruning (for tree-based models) should be considered to mitigate overfitting and improve generalization.

Another contributing factor to classification difficulty is the potential overlap in feature space between the "LAYAK" and "BELUM LAYAK" classes. The reduced discriminative performance, especially in Folds 2 and 3, suggests that both classes share similar characteristics across key variables. This overlap underscores the importance of more sophisticated feature engineering and selection processes. Incorporating domain knowledge, applying dimensionality reduction techniques, or leveraging feature importance analysis may help identify more discriminative attributes and reduce class ambiguity.

To validate the robustness of the classification results, statistical significance testing was conducted using McNemar's [19] Test on the best-performing fold. The resulting p-value ($p < 0.05$) indicates that the model's predictive performance is statistically significantly different from that of a random classifier. This finding confirms that the selected features, including economic status and academic performance indicators such as GPA, contribute meaningful predictive information rather than random variation. Although direct paired comparisons with prior studies were limited due to the unavailability of raw prediction outputs, the consistency of F1-score values across all cross-validation folds serves as an empirical proxy for model reliability. Future work should incorporate a formal McNemar's Test on an independent hold-out test set to rigorously compare the proposed hybrid clustering-classification framework against conventional single-model approaches.

Table 5. Comparison Statistic Performance Model

Metric	Mean Score	Std. Deviation (σ)	95% Confidence Interval
Accuracy	90.61%	±1.2%	[89.4%, 91.8%]
F1-Score	0.89	±0.02	[0.87, 0.91]
Precision	0.91	±0.01	[0.90, 0.92]

Source: (Research Results, 2025)

Overall, the model demonstrates strong generalization capability, as evidenced by an average training accuracy of 94.17% and a mean validation accuracy of 90.61%. The relatively small gap between training and validation performance, supported by a low standard deviation ($\sigma \approx \pm 1.2\%$), indicates stable and consistent behavior across folds. This suggests that the model is not only accurate but also robust, providing reliable predictive performance in the context of educational data mining.

CONCLUSION

The results of the model evaluation across five folds demonstrated consistent and robust performance, with accuracy ranging from 89% to 97%. Despite the strong performance within the five folds, it is important to note that the model's generalization capabilities are based solely on internal 5-fold cross-validation of 300 samples. No independent test set or external dataset was used, which limits the external validity of the model. Consequently, the model's ability to perform in real-world scenarios or with datasets beyond the 300 samples in this study remains uncertain. This limitation suggests that the findings might not fully extend to other student populations or contexts. Additionally, while fairness [20] and ethical considerations are discussed in the literature review, the study did include specific fairness metrics or subgroup analyses. Given the socio-economic implications of scholarship allocation, addressing these concerns is essential. Failure to incorporate fairness assessments could potentially introduce bias in the model, particularly against certain student groups.

Future research should focus on implementing fairness metrics and external validation to enhance the model's reliability, applicability, and equity in real-world applications. This approach would ensure that scholarships are awarded to the most deserving candidates based on academic performance, leadership potential, and financial need, without unintended biases. Table 2 presents a unified comparison between binary and multi-class classification performance. The results show that while XGBoost slightly outperforms other models in binary accuracy, Random Forest demonstrates more stable performance across both binary and multi-class settings. As expected, performance decreases in the multi-class scenario due to increased classification complexity. Class-wise evaluation (Table 3) indicates that the "Not Eligible" class achieves the highest performance, while confusion is more prominent between "Prestasi" and "Ashabunnajah," suggesting overlapping feature characteristics.

REFERENCE

[1] E. Gazali and A. A. Budiana, "A Bibliometric Analysis of Pesantren's Educational Impact: Insights from The Scopus Database (1994–2022)," Jurnal Pendidikan Islam, vol. 12, no. 1, pp. 15–33, 2023, doi: 10.14421/jpi.2023.121.15-33.

[2] K. M. Lestari, S. Zakir, D. Ilmi, and R. A. Gusli, "Evaluasi perubahan Kurikulum 2013 dengan Kurikulum Merdeka di SMAN 3 Bukittinggi," Idarah Tarbawiyah: Journal of Management in Islamic Education, vol. 5, no. 2, 2024, doi: 10.32832/itjmie.v5i2.16620.

[3] O. W. P. T. L. Sofian Lusa, Peran e-Commerce dalam Mendukung Ekonomi Digital Indonesia, Yogyakarta, Indonesia: Andi Offset, 2024.

[4] T. Lestari, "Ethics in Technological Innovation: Strengthening Human Responsibility and Values," in Proceeding of International Conference on Islamic Boarding School, vol. 1, no. 1, pp. 138–144, 2025, doi: 10.61159/icop.v1i1.428.

[5] H. Alias, M. Adif, M. A. Abdul Aziz, N. Hambali, and M. N. Taib, "Student performance classification: A comparison of feature selection methods based on online learning activities," International Journal of Electrical and Computer Engineering, vol. 14, no. 4, pp. 4675–4685, 2024, doi: 10.11591/ijece.v14i4.pp4675-4685.

[6] A. R. Tutik Lestari, "Transformation of Pesantren Education in the Digital Era: AI Innovation and Adaptation for Technology-Based Learning," *The Electronic Integrated Computer Algorithm Journal,* vol. 2, no. 1, pp. 69-90, 2025, doi: 10.62123/enigma.v2i2.58.

[7] U. Kannengiesser and J. S. Gero, "Modelling the Design of Models: An Example Using CRISP-DM," Proceedings of the Design Society, vol. 3, pp. 2705–2714, 2023, doi: 10.1017/pds.2023.271.

Accredited Rank 2 (Sinta 2) based on the Decree of the Dirjen Penguatan RisBang Kemenristekdikti No.225/E/KPT/2022, December 07, 2022. Published by LPPM Universitas Nusa Mandiri

VOL. 11. NO. 4 MAY 2026
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v11i4.7196

[8] M. A. Karabiyik, B. Turkoglu, and T. Asuroglu, "A cluster-assisted differential evolution-based hybrid oversampling method for imbalanced datasets," PeerJ Computer Science, vol. 11, p. e3177, 2025, doi: 10.7717/peerj-cs.3177.

[9] S. Nazuah, S. S. Hilabi, A. Hananto, B. Huda, and Tukino, "Seleksi Penerimaan Beasiswa Dengan Metode K-Means Clustering Menggunakan Orange," *JUSTINDO: Jurnal Sistem dan Teknologi Informasi Indonesia,* vol. 8, no. 1, p. 1–10, 2023, doi: 10.32528/justindo.v8i1.212.

[10] R. A. Nugrahaeni and K. Mutijarsa, "Comparative analysis of machine learning KNN, SVM, and random forests algorithm for facial expression classification," in 2016 International Seminar on Application for Technology of Information and Communication (ISemantic), 2016, pp. 163–168, doi: 10.1109/ISEMANTIC.2016.7873831.

[11] M. Kumar, N. Singh, J. Wadhwa, and P. Singh, "Utilizing Random Forest and XGBoost Data Mining Algorithms for Anticipating Students' Academic Performance," I.J. Modern Education and Computer Science, vol. 16, no. 2, pp. 29–44, 2024, doi: 10.5815/ijmecs.2024.02.03.

[12] E. Cahapin, B. Malabag, C. Santiago Jr., J. Reyes, G. Legaspi, and K. Adrales, "Clustering of students admission data using k-means, hierarchical, and DBSCAN algorithms," Bulletin of Electrical Engineering and Informatics, vol. 12, no. 6, pp. 3647–3656, 2023, doi: 10.11591/eei.v12i6.4849.

[13] M. Sompa and R. Ishak, "Clustering Tingkat Ekonomi Mahasiswa Calon Penerima Kartu Indonesia Pintar (KIP) Kuliah Metode K-Means," Jurnal Ilmiah Ilmu Komputer Banthayo Lo Komputer, vol. 1, no. 2, pp. 65–71, 2022, doi: 10.37195/balok.v1i2.175.

[14] V. W. Lumumba, D. Kiprotich, M. L. Mpaine, N. G. Makena, and M. D. Kavita, "Comparative Analysis of Cross-Validation Techniques: LOOCV, K-folds Cross-Validation, and Repeated K-folds Cross-Validation in Machine Learning Models," SSRN Electronic Journal, Jun. 2024, doi: 10.2139/ssrn.5266507.

[15] S. Romero, X. Li, N. Xi, R. A. Romero, and M. S.-V. Romero, "Statistical and machine learning models for predicting university dropout and scholarship impact," PLOS ONE, vol. 20, no. 6, p. e0325047, 2025, doi: 10.1371/journal.pone.0325047.

[16] B. Zhu, X. Jing, L. Qiu, and R. Li, "An Imbalanced Data Classification Method Based on Hybrid Resampling and Fine Cost Sensitive Support Vector Machine," Computers, Materials & Continua, vol. 79, no. 3, p. 3977, 2024, doi: 10.32604/cmc.2024.048062.

[17] Y. Zhang, L. Deng, and B. Wei, "Imbalanced Data Classification Based on Improved Random-SMOTE and Feature Standard Deviation," Mathematics, vol. 12, no. 11, p. 1709, 2024, doi: 10.3390/math12111709.

[18] M. Han, A. Li, Z. Gao, D. Mu, and S. Liu, "Hybrid Sampling and Dynamic Weighting-Based Classification Method for Multi-Class Imbalanced Data Stream," Applied Sciences, vol. 13, no. 10, p. 5924, 2023, doi: 10.3390/app13105924.

[19] C. S. Metzner, S. Gao, D. Herrmannova, E. Lima-Walton, and H. A. Hanson, "Attention Mechanisms in Clinical Text Classification: A Comparative Evaluation," IEEE Journal of Biomedical and Health Informatics, vol. 28, no. 4, pp. 2247–2258, 2024, doi: 10.1109/JBHI.2024.3355951.

[20] J. Yang, A. A. S. Soltan, D. W. Eyre, et al., "Algorithmic fairness and bias mitigation for clinical machine learning with deep reinforcement learning," Nature Machine Intelligence, vol. 5, pp. 884–894, 2023, doi: 10.1038/s42256-023-00697-3.