Vol. 8, No. 4, August 2026

ISI Journal of Information Systems and Informatics

Published By Asosiasi Doktor Sistem Informasi Indonesia

Machine Learning-Based Multi-Class Scholarship Classification with a Streamlit Prototype

Siti Nabila Ariza Fitri[1], Rona Nisa Sofia Amriza[2], M. Yoka Fathoni[3]

[1,2,3] Information Systems, Telkom University, Purwokerto Campus, Indonesia

Received:
February 17, 2026
Revised:
May 27, 2026
Accepted:
July 29 2026
Published:
August 22, 2026

Corresponding Author:
Author Name*:
Rona Nisa Sofia Amriza
Email*:
ronanisa@telkomuniversity.ac.id

DOI:
10.63158/journalisi.v8i4.1673

Abstract. The scholarship selection process at a private university is still conducted manually by reviewing scholarship applicant documents individually, resulting in a time-consuming process and potential inconsistencies in evaluation. This study aims to develop a multi-class Machine Learning-based classification model to support scholarship classification based on recipient criteria and evaluating model performance using accuracy, precision, recall, F1-score, and confusion matrix metrics. This study applied the Cross-Industry Standard Process for Data Mining (CRISP-DM) framework consisting of six stages Business Understanding, Data Understanding, Data Preparation, Modelling, Evaluation, and Deployment. The dataset consists of 408 historical scholarship recipient records categorized into four scholarship classes. A key contribution of this study is the integration of the CRISP-DM methodology with feature importance analysis using Random Forest Feature Importances, class balancing using SMOTE, and machine learning classifiers, including Random Forest, SVM, and Naïve Bayes, within a unified predictive modelling framework. Based on the evaluation results, the Support Vector Machine algorithm achieved the best performance with an accuracy of 77% and a macro F1-score of 63%, followed by Random Forest at 76% and Naïve Bayes at 70%. The SVM model was then implemented as a Streamlit-based web to support scholarship recommendations and is not intended as a final scholarship approval system. This research contributes to the development of an efficient, data-driven scholarship classification support system for higher education institutions.

Keywords: Scholarship Classification; Multi-Class Classification; Machine Learning Algorithms; CRISP-DM; Streamlit.

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

1. INTRODUCTION

Education plays an important role in developing individual potential and improving the quality of human resources to compete in the era of globalization [1]. One of the efforts made by universities to support equal access to education is through scholarship programs [2]. Scholarships are designed to recognize students' academic and non-academic achievements while encouraging them to maintain and improve their performance [3]. In addition, scholarships are also in line with the fourth Sustainable Development Goals (SDGs), namely Quality Education, which aims to ensure access to quality education for all Indonesian citizens [4], [5]

The University generally offers several scholarship programs with different objectives and selection criteria. in this study, four scholarship categories were considered: Full Scholarships, KIP, OPES, and Endowment Fund. Each type of scholarship has different criteria, including academic achievement, non-academic achievements, and students' economic circumstances. Therefore, selecting the most appropriate scholarship category requires careful evaluation, as each program has different eligibility criteria.

The university currently implements a manual document selection process by checking each applicant's document one by one [6]. This process is time-consuming as the number of scholarship applicants increases. Therefore, this study aims to reduce the manual workload by developing a machine learning model to classify scholarship types based on recipient criteria. This process requires a machine learning approach to help model and potentially improves efficiency and consistency in the scholarship management process. This process requires a Machine Learning approach is needed to help model and automate the selection process. This approach allows the system to learn from training data, thereby automatically recognizing patterns and classifying scholarship recipients more accurately [7]

Education Data mining (EDM) is a field of research that applies data mining and machine learning techniques to analyze educational data and support data-driven decision-making in educational settings [8]. The application of machine learning techniques in scholarship selection can be considered part of EDM because it utilizes student data to support objective and consistent decision-making processes.

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

Machine Learning is a part of Artificial Intelligence that can learn automatically from the data provides and can improve its performance without having to be explicitly taught [9]. Research by Niarman and Adli using data from 1,048 scholarship recipients showed that the best algorithm is Random Forest, which produces the most accurate model at 83%. followed by Decision Tree with an accuracy of 72%, Logistic Regression of 69%, and the lowest accuracy in the Support Vector Machine of 68% [10]. Another research utilized 270 data points from KIP scholarship recipient students and applied the Naïve Bayes algorithms. Evaluation was conducted using accuracy, precision, recall, and F1-score, yielding an accuracy of 98.5%, and an F1-score of 98.5%, indicating that the probabilistic approach also performs well in scholarship decision support systems [11].

Although previous studies have shown good results in predicting scholarship recipients, several challenges remain. Many existing studies focus on binary classification or investigate only a single scholarship program, whereas universities often manage multiple scholarship schemes simultaneously. Furthermore, relatively few studies account for class imbalance, even though scholarship datasets typically contain unequal numbers of records across different scholarship categories. Furthermore, the practical implementation of machine learning models through web-based applications has received insufficient attention, thereby limiting their application in supporting real-world administrative processes.

This study addresses these limitations by developing a multi-class scholarship classification model using historical scholarship recipient data obtained from a private university. This dataset consists of four scholarship categories: Full Scholarship, KIP, OPES, and Endowment Fund. To mitigate the effects of class imbalance, the Synthetic Minority Over-sampling (SMOTE) technique was applied to the training data prior to model development. Three classification algorithms: Random Forest, Support Vector Machine, and Naïve Bayes. Implemented following the CRISP-DM methodology. Additionally, feature importance analysis was conducted to identify the variables that most significantly contribute to the classification process.

This study makes three contributions. First, the proposed approach addresses the problem of multi-class scholarship classification involving four scholarship categories by using historical records of scholarship recipients. Second, this study evaluates and

compares the performance of Random Forest, Support Vector Machine, and Naïve Bayes on an imbalanced scholarship dataset after applying SMOTE. Finally, the model with the best performance is implemented as a Streamlit-based prototype that provides classification support for scholarship administration. This prototype is intended to assist in the scholarship evaluation process and should not be interpreted as an automated decision-making system or a substitute for institutional scholarship policies. Therefore, this study aims to develop and evaluate machine learning models for multi-class scholarship classification using historical scholarship recipient records. The performance of the proposed models is evaluated using accuracy, precision, recall, and F1-score. Additionally, the best-performing model is implemented as a web-based prototype developed to assist administrators in the classification process and is not intended to replace policies or final decisions regarding the selection of scholarship recipients.

2. METHODS

This research was conducted based on the Cross-Industry Standard Process for Data Mining (CRISP-DM) methodology framework to systematically organize the research process. starting from Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation, and Deployment [12] [13]. This framework is extensively applied in various analytical and data science studies due to its structure and practical approach [14]

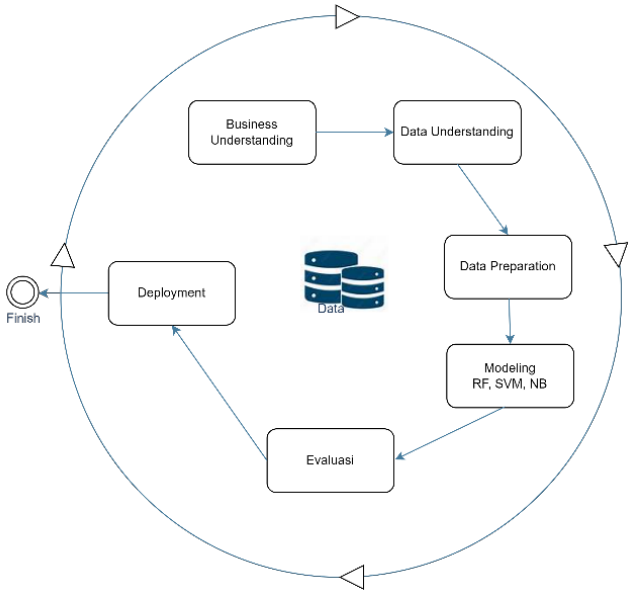

Figure 1. Research Flow

2.1. Business Understanding

The University has a target of enrolling approximately 1,600 new students. Each type of scholarship managed does not have a specific quota target for the number of recipients. Scholarship recipients are determined based on budget availability.

2.2. Data Understanding

The dataset used in this study is secondary data in the form of historical scholarship recipient records obtained from the Student Affairs Unit and New Student Selection at a private university. The dataset was used solely for research purposes with permission from the university responsible unit. The dataset was anonymized prior to analysis and does not contain any personally identifiable information, such as students' names, ID numbers, addresses, or other sensitive personal data.

The dataset consists of 408 records of confirmed scholarship recipients categorized into four scholarship classes: Full Scholarship (Label 0), Endowment Fund (Label 1), KIP (Label 2) and OPES (Label 3). The original dataset contains eleven attributes: Study Program, High School Origin, GPA, Academic achievement, non-Academic achievement, report card score, exam test score, economic conditions, letter of financial hardship, rank from 1 to 15, and Telkom School of Origin.

2.3. Data Preparation

Before the modelling phase, the dataset undergoes several pre-processing steps to improve data quality and ensure it is ready for analysis. These steps involved first, data cleaning, which removed two attributes: school of origin and program. Both attributes were removed because they were not directly related to any scholarship eligibility criteria. After data cleaning, the dataset retained nine features for modeling. Second data transformation, categorical encoding, was applied to convert text-based features into numeric values. Third, splitting data into training and testing sets in an 80:20 ration. Fourth, Feature importance. Fifth, the Synthetic Minority Oversampling (SMOTE) technique is applied exclusively to the training set after the division to prevent data leakage. This pre-processing stage aims to improve the performance and stability of the classification model [15].

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

2.4. Modelling

The modelling process focuses on developing classifications that can distinguish scholarship categories based on the characteristics of prospective recipients. In this study, three machine learning algorithms were employed in this study, namely Random Forest, Support Vector Machine, and Naïve Bayes. The Random Forest model was configured with 50 decision trees (n_estimators = 50) and a random state of 42. The SVM model used a linear kernel with a regularization parameter (C) of 1. Meanwhile, the Naïve Bayes classifier utilized the Gaussian Naïve Bayes variant.

The implementation phase was implemented through a web system developed using the Streamlit Framework. The experimental environment used in this study consisted of Python version 3.12.13 as the primary programming language. Machine learning model development and evaluation were performed using Scikit-learn version 1.6.1, while class imbalance handling was performed using Imbalanced-learn version 0.14.1 via the SMOTE technique. SMOTE was applied exclusively to the training set after the split to address class imbalance. A web-based scholarship recommendation was developed using Streamlit version 1.56.0 and implemented via Visual Studio Code (VS Code).

Random Forest is a classification algorithm with a supervised learning type that uses a tree branching method or Decision Tre [16] [17][18]. The following are equations in data class decision making:

$$y(\times) = mode\ \{\mathrm{C}_1\}(\times), \dots, \mathrm{C}_2\ (\times), .., \mathrm{C}_m(\times) \quad (1)$$

Where m is the number of decision trees in the random forest and mode is the operator for selecting the input number with the highest frequency. The following is the equation for selecting data class decisions.

Support Vector Machine is a machine learning algorithm that uses the principle of structure risk minimization (SRM) to predict hyperplanes in data classification [19][20]. The following decision function used in support vector machine classification is expressed as:

$$\mathrm{y}(\times)\ = \mathrm{sign}\ (\mathrm{w}.\times\ + \mathrm{b}) \quad (2)$$

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

Where f (x) is the decision function that assigns a class label to instance x;w is the weight of vector; x is the feature vector of instance; b is the bias.

Naïve Bayes is a term related to simple probabilistic classification from an ancient theorem discovered by Thomas Bayes. It assumes independence between features and calculates posterior probabilities to determine the most likely class [21]. The Bayes Theorem formula is expressed as:

$$\mathrm{P(Y|X)} = \frac{P(X|Y)P(Y)}{P(X)} \tag{3}$$

Where:

P(Y|X) is the posterior probability of class Y given features X,

P(X|Y) is the likelihood of features X given class Y,

P(Y) is the prior probability of class Y,

P(X) is the marginal probability of features X.

2.5. Evaluation

Model performance was assessed through several quantitative metrics of each classification algorithm using accuracy, precision, recall, F1-score, and confusion matrix. These evaluation metrics are used to compare the effectiveness of each model. Accuracy is a value that describes how accurate the model is in classification [22]

$$\frac{TP + TN}{TP + TN + FP + FN} \tag{4}$$

Precision is a value that expresses the ratio between the requested data and the prediction results provided by the model [23]

$$\frac{TP}{TP + FP} \tag{5}$$

Recall or Sensitivity is a value that describes the completeness or success of the model in retrieving information [23]

$$\frac{TP}{TP + FN} \tag{6}$$

F1-Score is a comparison of the weighted average of precision and recall [23]

$$F1 - Score = 2 \times \frac{precision + recall}{recision + recall} \quad (7)$$

2.6. Deployment

Deployment is the final stage of this research process. Machine learning is used to implement a classification model in the form of a scholarship recommendation website. The system was built using the Streamlit framework, an open-source Python framework. The website for this research was created in a simple format using Visual Studio Code.

3. RESULTS AND DISCUSSION

3.1. Data Understanding

The dataset used in this study contains 408 records of scholarship recipients. Each record represents an applicant profile containing academic, demographic, and socioeconomic information used to determine the appropriate scholarship category. The dataset consists of eleven input attributes and one target variable, namely the scholarship type. The target variable represents the scholarship category assigned to each recipient, which becomes the classification output in this study. The scholarship categories consist of four classes: Full Scholarship (0), Endowment Fund Scholarship (1), KIP Scholarship (2), and OPES Scholarship (3). The distribution of each class in the original dataset is presented in Table 1.

Table 1. Class Distribution of the Original Dataset

Scholarship Type	Class Label	Number of Records
Full Scholarship	0	158
Endowment Fund Scholarship	1	43
KIP Scholarship	2	144
OPES Scholarship	3	63
Total		**408**

Table 1 illustrates the distribution of scholarship classes in the initial dataset. The scholarship type attribute was selected as the target variable because the main objective

of this study is to develop a classification model capable of recommending the most appropriate scholarship category for prospective recipients based on their academic performance, school background, and socioeconomic conditions. The dataset contains various attributes representing different aspects of recipient eligibility, including academic achievement, economic conditions, and supporting administrative criteria. Table 2 presents the first five records of the dataset.

Table 2. Samples of the Dataset

Type	Study Program	GPA	High School Origin	Academic Achievement	Non-Academic Achievement	Report Card Score	Exam Test Score	Economic Condition	Letter of Financial Rank (1–15)	Telkom School Origin
Full	Telecommunication Engineering	3.95	SMAN 3 TEGAL	1	2	89	92	Medium	0	1	No
KIP	Telecommunication Engineering	3.89	SMAS Muhammadiyah 2 Gemolong	2	3	92	95	Low	1	1	No
KIP	Telecommunication Engineering	3.88	SMKN 1 Karanganyar	4	2	89	90	Low	1	1	No
Full	Telecommunication Engineering	3.88	SMKN 2 Purwokerto	3	2	95	97	Medium	0	1	No
Full	Telecommunication Engineering	4.00	SMKN 1 Purwokerto	1	2	97	96	High	0	1	No

3.2. Data Preparation

The data preparation stage aims to transform the raw dataset into a suitable format for the modelling process. This stage consists of several procedures, including data cleaning, feature selection, data transformation, label encoding, dataset splitting, feature importance analysis, and handling class imbalance using the Synthetic Minority Oversampling Technique (SMOTE).

The first step in data preparation is data cleaning, which focuses on identifying and handling missing values and irrelevant attributes. Based on the data quality assessment presented in Table 3, the original dataset contained eleven attributes, with two missing values identified in the High School Origin attribute. Since incomplete school-origin information could introduce uncertainty and affect the classification performance, this attribute was removed from the dataset. In addition, the Study Program attribute was excluded because all records belonged to the same study program, resulting in no

discriminative contribution to the classification task. Removing these attributes reduces unnecessary complexity and ensures that only relevant features are retained for model development. After the cleaning process, the dataset consisted of nine predictive attributes, as presented in Table 3. These remaining attributes were selected because they represent relevant academic, socioeconomic, and supporting criteria associated with scholarship eligibility. The cleaned dataset was subsequently used as the input for the next preprocessing stages.

Table 3. Feature after cleaning

No	Attribute	Missing Value	Data Type
1	GPA	0	Float
2	Academic Achievement	0	Integer
3	Non-Academic Achievement	0	Integer
4	Report Card Score	0	Integer
5	Exam Test Score	0	Integer
6	Economic Condition	0	Integer
7	Letter of Financial Hardship	0	Integer
8	Rank from 1 to 15	0	Integer
9	Telkom School Origin	0	Integer

The next step is data transformation, where categorical variables are converted into numerical representations to ensure compatibility with machine learning algorithms. Two categorical attributes, namely Economic Condition and Telkom School Origin, required encoding. The Economic Condition attribute originally consisted of three categories: Low, Medium, and High. These categories were transformed into numerical values of 0, 1, and 2, respectively. Meanwhile, the Telkom School Origin attribute was encoded using binary representation, where No was converted into 0 and Yes was converted into 1. The target variable, namely the scholarship type, was also transformed through label encoding. Figure 2 presents the distribution of scholarship categories before encoding, where the target classes were still represented using categorical labels: Full Scholarship, KIP, OPES, and Endowment Fund Scholarship.

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

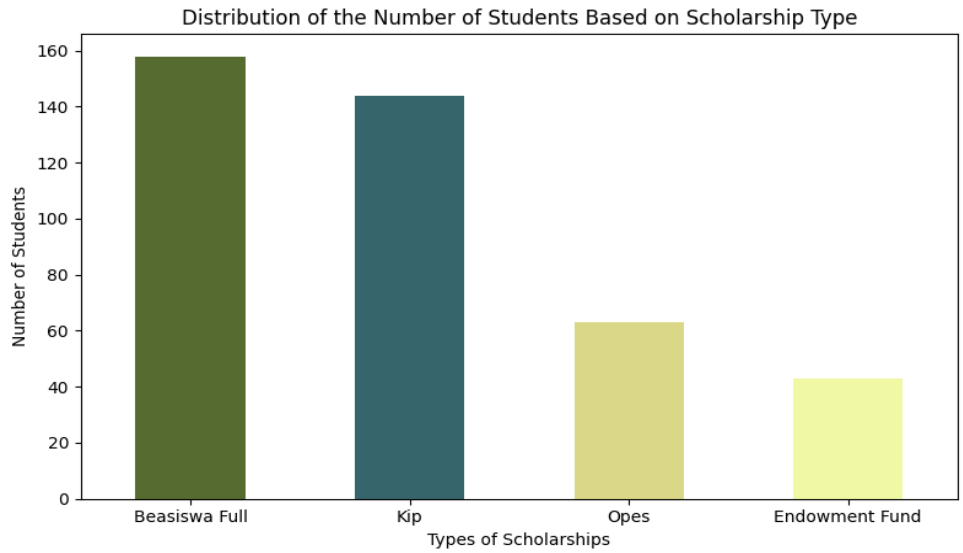

Figure 2. Visualization Distribution Class Before Encoding

After the encoding process, the scholarship categories were converted into numerical labels to enable classification modelling. The transformation process assigned numerical values to each scholarship category, where Full Scholarship became class 0, Endowment Fund Scholarship became class 1, KIP Scholarship became class 2, and OPES Scholarship became class 3. The distribution of the encoded target classes is illustrated in Figure 3.

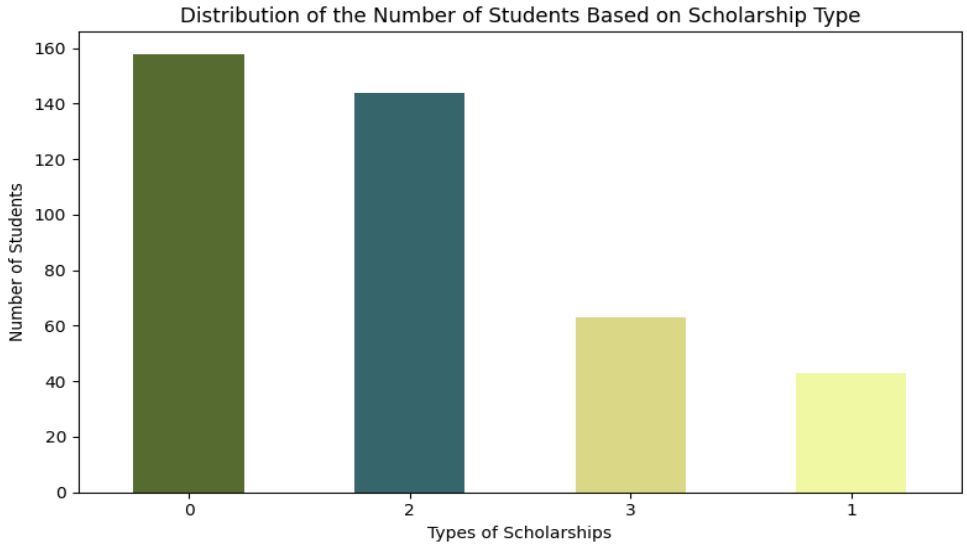

Figure 3. Visualization Distribution Class After Encoding

The next stage is dataset splitting, where the prepared dataset was divided into training and testing subsets. The splitting process was performed using an 80:20 ratio, where

80% of the data was allocated for model training and 20% was reserved for model evaluation. The distribution of the dataset after splitting is presented in Table 4.

Table 4. Splitting Data

Dataset	Amount	Percentage
Training	326	80%
Testing	82	20%

The training dataset consisted of 326 records distributed across four scholarship classes. The class distribution of the training subset is presented in Table 5.

Table 5. Class distribution data training after Splitting

Scholarship	Original Dataset
Full (0)	126
Endowment Fund (1)	35
KIP (2)	115
OPES (3)	50
Total	326

As shown in Table 5, the training dataset demonstrates an imbalanced class distribution, where the Endowment Fund Scholarship class contains significantly fewer samples compared with the Full Scholarship and KIP Scholarship classes. This imbalance may affect the ability of classification algorithms to learn minority classes effectively. The testing dataset contained 82 records, with the distribution of each scholarship class presented in Table 6.

Table 6. Class distribution data test after Splitting

Scholarship	Original Dataset
Full (0)	32
Endowment Fund (1)	8
KIP (2)	29
OPES (3)	13
Total	82

Feature importance analysis was then conducted to identify the contribution of each attribute to the classification process. This analysis was performed using the Random Forest algorithm to measure the relative importance of each feature. The visualization results are shown in Figure 4.

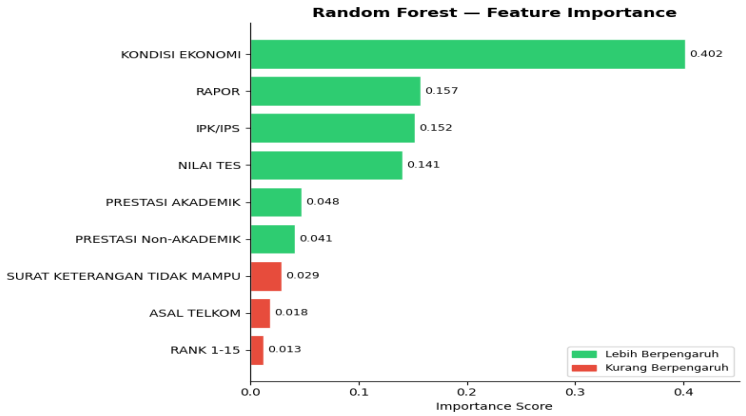

Figure 4. Features Importance

Based on Figure 4, the features with the highest contribution to scholarship classification were Economic Condition, Report Card Score, GPA, Exam Test Score, and Academic Achievement. These attributes represent the strongest factors influencing scholarship category prediction. Although some features demonstrated lower importance values, they were retained because each attribute represents a predefined scholarship selection criterion and may still contribute collectively during the modelling process.

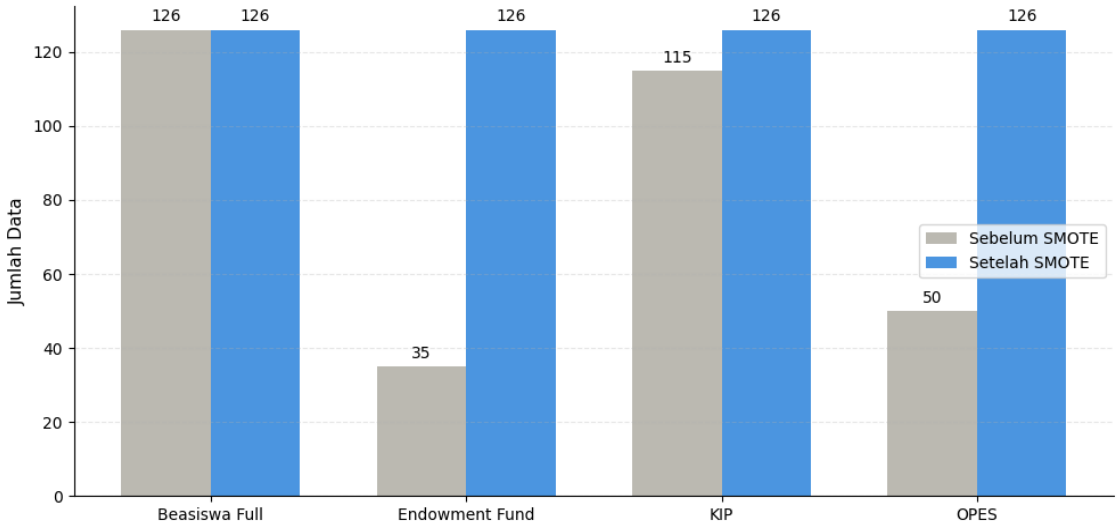

Figure 5. Visualization before-after SMOTE

The final step in the data preparation stage is handling class imbalance using Synthetic Minority Oversampling Technique (SMOTE). SMOTE was applied only to the training dataset after the splitting process to prevent information leakage between training and testing data. Figure 5 illustrates the comparison of class distribution before and after applying SMOTE.

As shown in Figure 5, the original training dataset contained an unequal distribution among scholarship classes. The minority class, Endowment Fund Scholarship, had substantially fewer samples than other categories. SMOTE generated synthetic samples for minority classes by interpolating existing samples, resulting in a more balanced training dataset. The detailed comparison before and after SMOTE is presented in Table 7.

Table 7. Dataset before-after SMOTE

Dataset Training	Before SMOTE	After SMOTE
Full	126	126
Endowment Fund	35	126
KIP	115	126
OPES	50	126
Dataset Training	326	504

Table 7 shows that SMOTE successfully balanced the training dataset by increasing the minority classes while maintaining the original majority class distribution. After applying SMOTE, each scholarship category contained 126 samples, increasing the total training data from 326 to 504 records. This balanced dataset was subsequently used for the modelling stage to improve the classifier's ability to recognize all scholarship categories.

3.3. Modelling and Performance Evaluation

Three supervised machine learning algorithms—Support Vector Machine (SVM), Random Forest (RF), and Naïve Bayes (NB)—were developed and evaluated for scholarship classification. The models classified applicants into four scholarship categories: Full Scholarship (Class 0), Endowment Fund (Class 1), KIP (Class 2), and OPES (Class 3). Model performance was assessed using accuracy, precision, recall, and F1-score. In addition to the hold-out test evaluation, 5-fold cross-validation was performed to assess the consistency and generalizability of the models across different subsets of the dataset.

3.3.1. Support Vector Machine Performance

As shown in Table 8, the SVM achieved an overall test accuracy of **77%**, with a weighted F1-score of **0.76** and a macro F1-score of **0.63**. The model demonstrated excellent performance for the Full Scholarship class, achieving a recall of 1.00 and an F1-score of 0.96. It also performed relatively well for KIP, with an F1-score of 0.75. However, performance for the Endowment Fund category was substantially weaker, with a precision of 0.20, recall of 0.12, and F1-score of 0.15. This result indicates that the model had difficulty distinguishing Endowment Fund applicants from applicants belonging to other scholarship categories.

Table 8. Classification report of the Support Vector Machine model

Class	Precision	Recall	F1-Score	Support
Full Scholarship (0)	0.91	1.00	0.96	32
Endowment Fund (1)	0.20	0.12	0.15	8
KIP (2)	0.86	0.66	0.75	29
OPES (3)	0.55	0.85	0.67	13
Accuracy			**0.77**	**82**
Macro average	**0.63**	**0.66**	**0.63**	**82**
Weighted average	**0.77**	**0.77**	**0.76**	**82**

The difference between the macro F1-score (0.63) and weighted F1-score (0.76) also indicates that model performance was uneven across the four classes. Since the weighted metric gives greater influence to classes with more observations, its higher value reflects the stronger performance achieved for the more represented classes, particularly Full Scholarship and KIP.

3.3.2. Random Forest Performance

Table 9 presents the performance of the Random Forest model. The model obtained an overall test accuracy of 76%, a macro F1-score of 0.67, and a weighted F1-score of 0.75. Similar to SVM, Random Forest correctly identified all Full Scholarship instances, resulting in a recall of 1.00 and an F1-score of 0.96. Random Forest showed substantially better performance than SVM for the Endowment Fund class, increasing the F1-score from 0.15 to 0.50. Its performance for KIP was comparatively balanced, with precision, recall, and F1-score all approximately 0.69. For OPES, however, the model achieved an F1-score of

only 0.52. Although the overall accuracy of Random Forest was slightly lower than that of SVM on the hold-out test set, its macro F1-score was higher (0.67 versus 0.63). This suggests that Random Forest provided a more balanced classification performance across the four scholarship categories.

Table 9. Classification report of the Random Forest model

Class	Precision	Recall	F1-Score	Support
Full Scholarship (0)	0.91	1.00	0.96	32
Endowment Fund (1)	0.75	0.38	0.50	8
KIP (2)	0.69	0.69	0.69	29
OPES (3)	0.50	0.54	0.52	13
Accuracy			**0.76**	**82**
Macro average	**0.71**	**0.65**	**0.67**	**82**
Weighted average	**0.75**	**0.76**	**0.75**	**82**

3.3.3. Naïve Bayes Performance

As presented in Table 10, Naïve Bayes obtained an overall test accuracy of 70%, which was lower than those of SVM and Random Forest. The model achieved a macro F1-score of 0.61 and a weighted F1-score of 0.70. Naïve Bayes correctly classified all Full Scholarship instances and achieved an F1-score of 0.96 for this class. Interestingly, the model obtained perfect precision (1.00) for KIP; however, its recall was only 0.38. This indicates that predictions assigned to KIP were highly reliable, but a substantial proportion of actual KIP applicants were classified into other categories. For OPES, the model achieved a recall of 0.77 and an F1-score of 0.61. Performance for Endowment Fund remained relatively weak, with an F1-score of 0.33. Overall, these findings show that the three algorithms exhibited different class-specific strengths. SVM produced the highest hold-out test accuracy, Random Forest achieved the highest macro F1-score, and Naïve Bayes demonstrated high precision but relatively low recall for the KIP category.

Table 10. Classification report of the Naïve Bayes model

Class	Precision	Recall	F1-Score	Support
Full Scholarship (0)	0.91	1.00	0.96	32
Endowment Fund (1)	0.25	0.50	0.33	8

Class	Precision	Recall	F1-Score	Support
KIP (2)	1.00	0.38	0.55	29
OPES (3)	0.50	0.77	0.61	13
Accuracy			**0.70**	**82**
Macro average	**0.67**	**0.66**	**0.61**	**82**
Weighted average	**0.81**	**0.70**	**0.70**	**82**

3.3.4. Cross-Validation Performance

To complement the hold-out test evaluation, a 5-fold cross-validation procedure was conducted to examine the stability of model performance across different data partitions.

Table 11. Five-fold cross-validation accuracy

Model	Mean Cross-Validation Accuracy
Support Vector Machine	69.84%
Random Forest	**76.57%**
Naïve Bayes	67.65%

The cross-validation results in Table 11 reveal that Random Forest achieved the highest mean accuracy of 76.57%, followed by SVM with 69.84% and Naïve Bayes with 67.65%. These results provide an important complement to the single hold-out test evaluation. Although SVM achieved the highest test-set accuracy (77%), its lower cross-validation accuracy suggests that its performance was more dependent on the particular train-test partition. In contrast, the Random Forest model maintained an average cross-validation accuracy of 76.57%, which was very close to its hold-out accuracy of 76%. This consistency indicates greater stability across different subsets of the data. Therefore, when both predictive accuracy and cross-validation stability are considered, Random Forest provides the strongest overall performance among the three evaluated algorithms.

3.3.5. Confusion Matrix Analysis

Figure 6 show the confusion matrix for the SVM model. Of the 82 test instances, 63 were correctly classified, corresponding to an accuracy of approximately 77%. All 32 Full Scholarship applicants were correctly identified. For the KIP and OPES categories, the

model correctly classified 19 of 29 and 11 of 13 instances, respectively. In contrast, only 1 of the 8 Endowment Fund applicants was correctly identified. The largest source of error for Endowment Fund was confusion with OPES, with five Endowment Fund instances being predicted as OPES. KIP applicants were also distributed among several incorrect categories, including Full Scholarship, Endowment Fund, and OPES. These errors suggest that the characteristics separating these scholarship categories are less distinct than those associated with the Full Scholarship category.

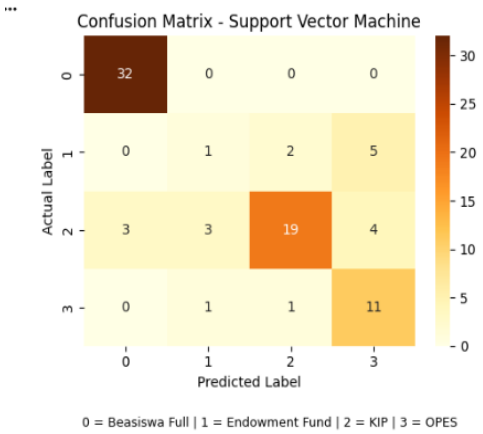

Figure 6. Confusion matrix of the Support Vector Machine model

Figure 7 and Table 13 present the confusion matrix for Random Forest. The model correctly classified 62 of the 82 test instances, corresponding to an accuracy of approximately 76%. All 32 Full Scholarship cases were classified correctly. In addition, the model correctly predicted 20 of 29 KIP cases, 7 of 13 OPES cases, and 3 of 8 Endowment Fund cases.

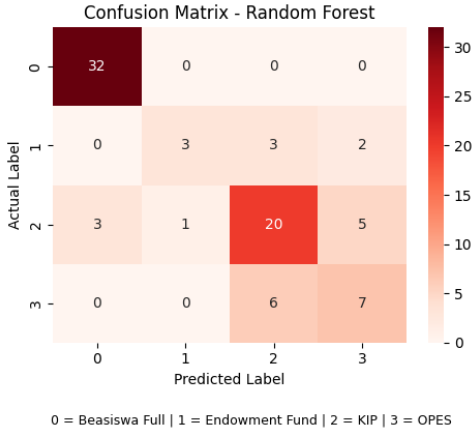

Figure 7. Confusion matrix of the Random Forest model

Compared with SVM, Random Forest substantially improved recognition of the Endowment Fund class, although some Endowment Fund applicants were still classified as KIP or OPES. A notable error pattern was observed between OPES and KIP, with 6 of the 13 OPES cases being classified as KIP. These results indicate that the two categories share feature patterns that are more difficult for the model to distinguish. Figure 8 and Table 14 show the classification errors generated by Naïve Bayes. The model correctly classified 57 of the 82 test instances, corresponding to an overall accuracy of approximately 70%. All Full Scholarship applicants were correctly identified, while 4 of 8 Endowment Fund applicants, 11 of 29 KIP applicants, and 10 of 13 OPES applicants were correctly classified.

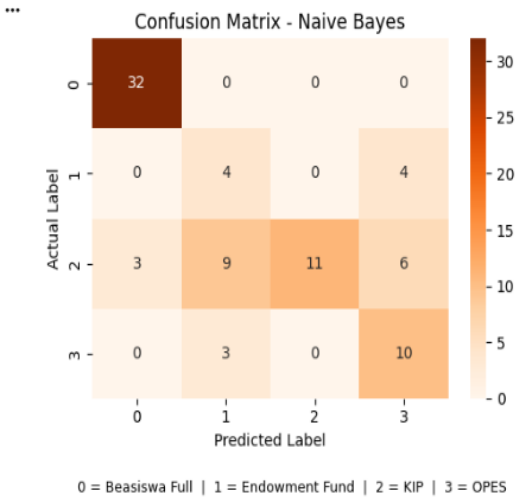

Figure 8. Confusion matrix of the Naïve Bayes model

The primary weakness of Naïve Bayes was its ability to recognize the KIP class. Although its precision for KIP was 1.00, only 11 of the 29 actual KIP applicants were correctly identified. Nine KIP instances were classified as Endowment Fund, six as OPES, and three as Full Scholarship. This explains the relatively low KIP recall of 0.38. In contrast, the model demonstrated relatively strong sensitivity to OPES, correctly detecting 10 of the 13 cases.

Across all three confusion matrices, the Full Scholarship category was consistently the easiest class to distinguish, as all models achieved a recall of 1.00 for this category. The largest classification difficulties occurred among Endowment Fund, KIP, and OPES. The unequal number of observations across classes may also contribute to these differences,

because the Endowment Fund category contained only eight test observations compared with 32 Full Scholarship and 29 KIP observations. Consequently, overall accuracy alone is insufficient for model selection, and macro-averaged metrics and cross-validation performance should also be considered.

3.3.6. Overall Model Comparison

The combined evaluation demonstrates that no single performance metric provides a complete assessment of model quality. SVM achieved the highest hold-out test accuracy (77%), while Random Forest achieved a slightly lower accuracy (76%) but the highest macro F1-score (0.67) and the highest 5-fold cross-validation accuracy (76.57%). Naïve Bayes obtained the lowest overall accuracy (70%) and cross-validation accuracy (67.65%).

The cross-validation findings are particularly important for model selection. The close agreement between the Random Forest hold-out accuracy (76%) and cross-validation accuracy (76.57%) indicates relatively stable performance across different data partitions. By comparison, SVM exhibited a larger difference between its hold-out accuracy (77%) and mean cross-validation accuracy (69.84%). Based on the combined evidence from accuracy, class-balanced performance, confusion matrices, and cross-validation, Random Forest was selected as the most suitable model for the scholarship recommendation system.

3.3.7. Proposed Machine Learning Architecture

Figure 9 illustrates the overall architecture of the proposed scholarship recommendation framework. The process begins with the scholarship applicant dataset, followed by data preprocessing to prepare the variables for machine learning analysis. Random Forest-based feature importance analysis is subsequently employed to identify the contribution of the input variables to scholarship classification. The processed data are then used to train and evaluate three classification algorithms: Support Vector Machine, Random Forest, and Naïve Bayes.

Model performance is evaluated using accuracy, precision, recall, F1-score, confusion matrices, and 5-fold cross-validation. This multi-metric evaluation enables the models to be assessed not only in terms of overall predictive accuracy but also in terms of class-specific performance and consistency across different data partitions. Based on the experimental results, Random Forest demonstrates the most consistent overall

performance and is therefore selected for implementation in the final system. The selected model is subsequently integrated into a Streamlit-based prototype to provide data-driven scholarship recommendations.

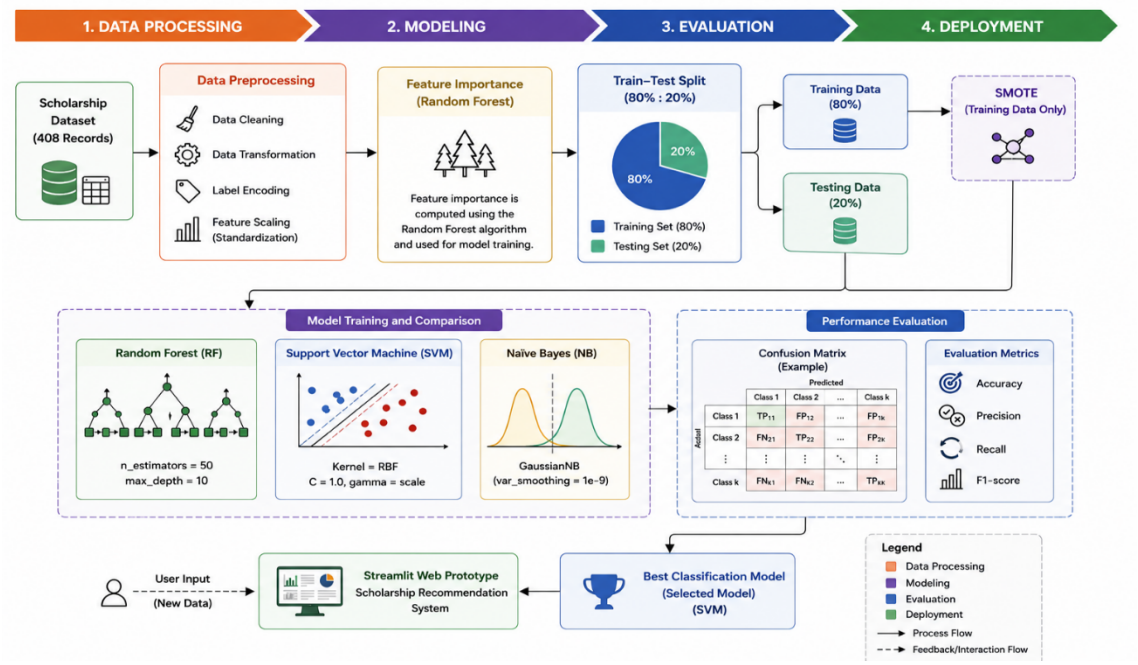

Figure 9. Architecture of the proposed machine learning-based scholarship recommendation system

3.4. Deployment

To demonstrate the practical applicability of the proposed scholarship classification approach, the selected Random Forest model was integrated into a web-based prototype developed using the Streamlit framework. The prototype translates the experimental machine learning model into an interactive decision-support interface that enables scholarship administrators to enter applicant information, review available scholarship categories, and obtain model-generated scholarship recommendations. The prototype is designed as a decision-support tool rather than an autonomous decision-making system. Its output provides an initial data-driven recommendation that may assist administrators in screening and evaluating scholarship applicants. Final scholarship decisions remain subject to institutional regulations, administrative verification, eligibility requirements, and review by authorized personnel. This design is intended to ensure that machine learning supports, rather than replaces, human judgment and established scholarship policies.

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

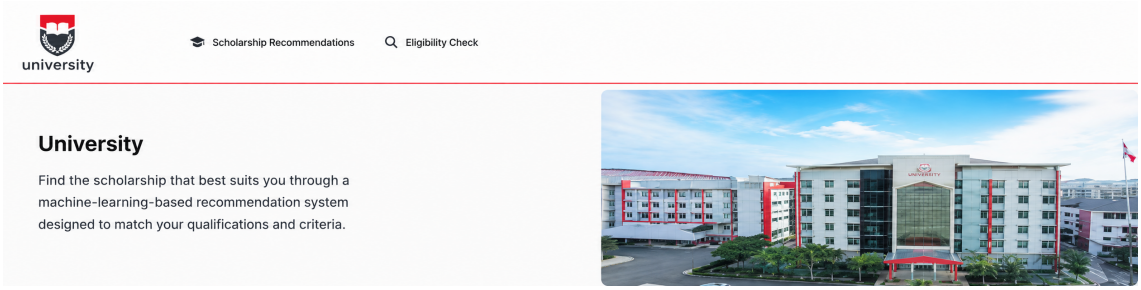

Figure 10. Website homepage of the scholarship recommendation prototype

Figure 10 presents the homepage of the developed scholarship recommendation prototype. The homepage provides general information about the purpose and functionality of the system and serves as the primary navigation interface. Through the navigation panel, users can access the main system functions, including scholarship information, scholarship recommendations, and applicant eligibility assessment. The interface was designed to provide a simple workflow so that users can move directly from general system information to the scholarship assessment process.

Figure 11 presents the Scholarship Information page, which provides an overview of the four scholarship categories considered in this study: Full Scholarship, Endowment Fund, KIP, and OPES. This page enables users to review the available scholarship alternatives before performing applicant classification. Presenting the scholarship categories within the prototype also provides contextual information that may assist administrators in interpreting the recommendation generated by the classification model.

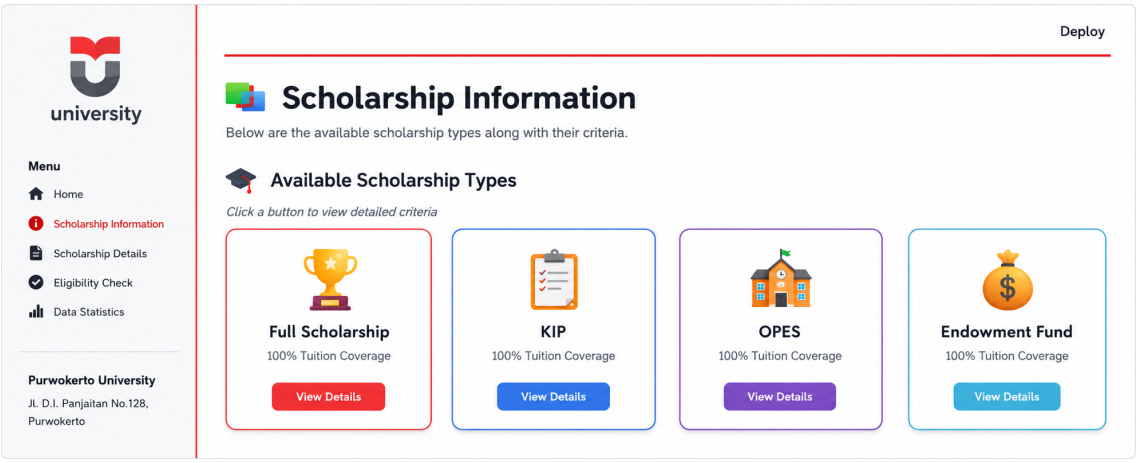

Figure 11. Scholarship Type Homepage

Figure 12 shows the scholarship eligibility assessment interface used to collect applicant information required by the classification model. Users enter the relevant applicant attributes, including academic achievement, non-academic achievement, economic conditions, and other variables included in the final predictive feature set. The submitted information undergoes the same preprocessing procedure used during model development before being passed to the trained Random Forest classifier.

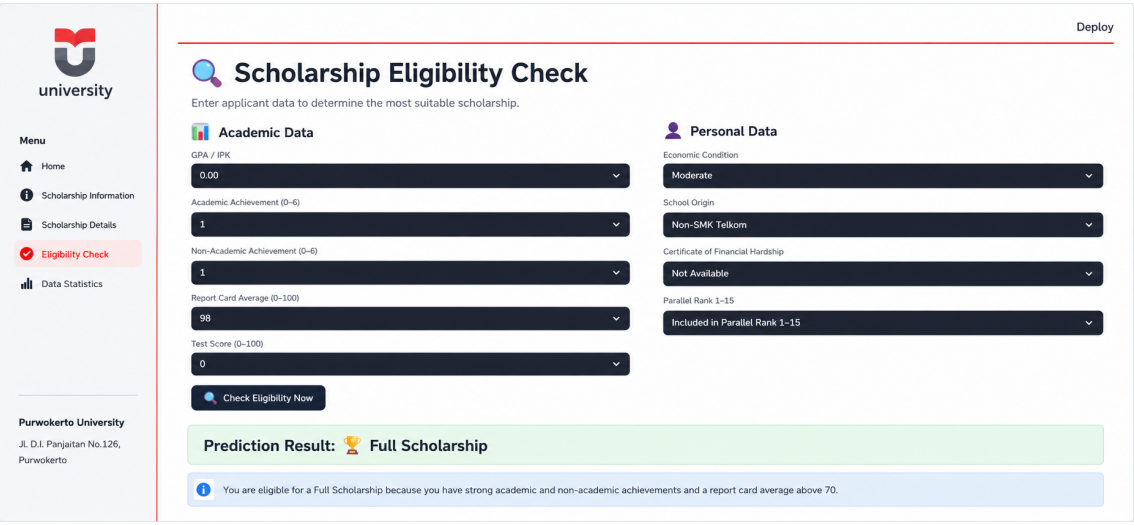

Figure 12. Eligibility Check Page

Based on these input attributes, the system generates a predicted scholarship category corresponding to one of the four classes: Full Scholarship, Endowment Fund, KIP, or OPES. The prediction should be interpreted as a model-based recommendation, rather than as an automatic confirmation of scholarship eligibility. Administrative staff are expected to verify the prediction against supporting documents and applicable institutional requirements before making a final decision.

The deployment stage demonstrates that the proposed classification model can be transformed from an experimental machine learning workflow into an operational prototype with a user-oriented interface. By integrating preprocessing, trained-model inference, and scholarship information within a Streamlit application, the prototype provides a practical mechanism for supporting the initial scholarship screening process while maintaining human oversight in the final decision-making procedure..

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

3.5. Discussion

The result showed that Support Vector Machine achieved the highest classification accuracy of 77%, followed by Random Forest at 76%, and Naïve Bayes at 70%. These results demonstrate Support Vector Machine superior performance compared to Random Forest and Naïve Bayes. However, the accuracy difference between Support Vector Machine and Random Forest is only 1%. From a practical perspective, this difference may not be significant, as both algorithms show comparable classification performance. Therefore, the selection of SVM as the best-performing model is not only based solely on accuracy but also considers its overall performance based on balanced set of evaluation metrics. Support Vector Machine is also capable of finding the optimal hyperplane for class separation, making it well-suited for multi-class scholarship classification with a limited number of features. However, in the 5-fold cross-validation evaluation, Random Forest achieved the highest average accuracy (76.57%), demonstrating more consistent performance across different data partitions, compared to SVM at 69.84% and Naive Bayes at 67.65%. This difference is due to the fact that cross-validation evaluates the model repeatedly on multiple data partitions, providing a more reliable estimate of the model's stability.

Although it achieved the highest accuracy, Support Vector Machine obtained a macro F1 score of 0.63, indicating that classification performance was not evenly distributed across all scholarship categories. This issue was particularly evident in the Endowment Fund class, which was one of the most difficult classes to classify because it has relatively fewer data points than the other classes. In the initial dataset, there were only 35 Endowment Fund scholarship recipients. This imbalance means the model has fewer examples to learn from for the Endowment Fund class. SMOTE was used to balance the training data, but the small number of test samples means that every classification error has a greater impact on the class's precision, recall, and F1-score. This condition increases the likelihood of overlap between classes, making it more difficult for the model to distinguish the Endowment Fund from other scholarship categories that also consider economic and academic achievement. Therefore, the classification performance for the Endowment Fund class remains lower compared to classes with larger data sets and clearer patterns.

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

Compared to previous research, the accuracy achieved by the Support Vector Machine this study is 77% higher than other study. Although previous studies [10] [11]. have shown good results in predicting scholarship, recipients most studies still focus on a single type of scholarship or classification with a limited number of categories. Furthermore, previous studies generally employ binary classification or a single scholarship scheme, whereas in practice, universities offer various scholarship programs with differing selection criteria. This situation highlights the need for a multiclass classification approach capable of classifying multiple types of scholarships simultaneously, thereby better representing the actual scholarship selection process.

The feature importance results revealed that "Economic Condition" was the most influential feature with an importance score of 0.403, ranking first in the Random Forest Feature Importances methods. This finding aligns with the scholarship criteria at a private university, where economic conditions serve as the primary differentiator between scholarship type KIP and Endowment Fund require low economic conditions, while the Full Scholarship prioritizes academic and non-academic achievements. However, this study utilized all available attributes, ensuring that classification decisions still consider a combination of factors according to the characteristics of each scholarship program. Focusing solely on economic conditions risks over-reliance on the economic condition attribute, thereby undermining the contribution of other attributes. Economic conditions did emerge as the most influential feature because it follows historical data patterns and the scholarship criteria applicable at a private university. Furthermore, the model continued to utilize other attributes, ensuring that classification was not determined solely by a single attribute.

This proposed model was developed using historical scholarship recipient records collected from a private university. As a result, the classification patterns learned reflect previous scholarship allocation decisions made based on the institution's scholarship criteria. Therefore, this proposed model serves as a classification aid. Because the model learns from historical data, its predictions may reflect historical allocation patterns and should always be interpreted in conjunction with the institution's scholarship regulations and verified by the administration to ensure fairness during the scholarship selection process.

Several limitations of this study should be acknowledged. The dataset consists of only 408 historical scholarship recipient records from a single institution, which may limit the generalizability of the proposed model to universities with different scholarship policies. Furthermore, although the SMOTE technique improved the balance of the training data, class imbalance persists in the test data because the original data distribution was preserved during evaluation. Future research is recommended to utilize a larger dataset, conduct external validation using data from various institutions, evaluate the model's fairness, integrate rule-based scholarship eligibility validation, and conduct usability testing involving scholarship administrators.

The best-performing model was implemented as a Streamlit-based web prototype to demonstrate its practical application in scholarship type classification. This prototype is intended to support scholarship administrators by providing classification results based on historical data of scholarship recipients. However, this prototype does not replace institutional scholarship regulations or final human decision-making. Instead, the prediction results are used as supporting information, and the final awarding of scholarships remains the responsibility of scholarship administrators in accordance with institutional policies.

4. CONCLUSION

This study developed a machine learning model for multi-class scholarship classification using historical records of scholarship recipients from a private university. The machine learning models were successfully built using three algorithms: Random Forest, Support Vector Machine, and Naive Bayes. Support Vector Machine achieved the highest performance on the hold-out test with an accuracy of 77.00%, while Random Forest achieved the highest cross-validation accuracy of 76.57%, demonstrating more consistent performance. The best-performing model was implemented as a Streamlit-based web prototype to assist scholarship administrators in classifying scholarship types, not to replace institutional regulations or final scholarship award decisions. Although the proposed approach shows promising performance, the classification of Endowment Fund categories remains limited due to the small number of training samples. Future research should consider larger and more diverse datasets, fairness evaluations, rule-based

validity checks, and usability testing by administrators to improve the practical implementation of the proposed prototype.

REFERENCES

[1] Y. Yunita, N. Hartono, and Erfina, "Penerapan data mining dalam sistem pendukung keputusan penerimaan beasiswa KIP-Kuliah menggunakan algoritma Naive Bayes," *J. Ilm. Sist. Inf. dan Ilmu Komput.*, vol. 5, no. 1, pp. 154–175, 2025, doi: 10.55606/juisik.v5i1.1415.

[2] H. Saleh and Hamria, "K-Nearest Neighbor berbasis seleksi atribut Chi Square untuk klasifikasi penerima beasiswa kurang mampu," *Simetris: J. Tek. Mesin, Elektro dan Ilmu Komput.*, vol. 14, no. 1, pp. 39–48, May 2023, doi: 10.24176/simet.v14i1.9178.

[3] B. R. A. Febrilia, M. H. Yani, and S. Anwar, "Ordinal logistic regression analysis of factors affecting students' scholarship status in Mataram University," *BAREKENG: J. Ilmu Mat. dan Terap.*, vol. 14, no. 2, pp. 227–232, Jun. 2020, doi: 10.30598/barekengvol14iss2pp227-232.

[4] S. Grobler, "Quality education in the context of the Sustainable Development Goals: An interpretation model," *Forum Oświatowe*, vol. 34, no. 1, pp. 153–164, 2022, doi: 10.34862/fo.2022.11.

[5] United Nations, "Goal 4: Quality education." Accessed: Aug. 22, 2026. [Online]. Available: United Nations Sustainable Development, Goal 4. Goal 4: Quality Education

[6] N. J. Saputra, "Penerapan algoritma K-Nearest Neighbor untuk klasifikasi penerima beasiswa pada STMIK Widya Cipta Dharma," B.S. thesis, Program Studi Teknik Informatika, STMIK Widya Cipta Dharma, Samarinda, Indonesia, 2024.

[7] M. Paramesha, S. S. Baba, S. F. Musthafa, V. V. Sai, and S. S. Kiran, "Data-driven scholarship prediction using machine learning for equitable education allocation," *Front. Collab. Res.*, vol. 2, no. 1S, pp. 43–50, Dec. 2024, doi: 10.70162/fcr/2024/v2/i1/v2i1s06.

[8] C. Romero and S. Ventura, "Educational data mining: A review of the state of the art," *IEEE Trans. Syst., Man, Cybern. C, Appl. Rev.*, vol. 40, no. 6, pp. 601–618, Nov. 2010, doi: 10.1109/TSMCC.2010.2053532.

[9] R. Y. Pradana, F. E. Nastiti, and I. Oktaviani, "Machine learning pengklasifikasikan performa karyawan direct sales force kartu prabayar menggunakan metode Random Forest classifier," *JEKIN: J. Tek. Inform.*, vol. 4, no. 3, pp. 590–599, Aug. 2024, doi: 10.58794/jekin.v4i3.864.

[10] A. Niarman, Asrida, and A. Adli, "Development of a machine learning-based predictive model for UPZ scholarship recipient selection at UIN Mahmud Yunus Batusangkar," *J. Edik Inform.*, vol. 11, no. 2, 2025, doi: 10.22202/ei.2025.v11i2.10022.

[11] D. B. Siswanto and D. Normawati, "Sistem klasifikasi monitoring dan evaluasi kelayakan penerima beasiswa UAD menggunakan algoritma Naïve Bayes," *J. Saintekom*, vol. 13, no. 2, pp. 161–172, Sep. 2023, doi: 10.33020/saintekom.v13i2.428.

[12] D. Ruswanti, D. Susilo, and R. Riani, "Implementasi CRISP-DM pada data mining untuk melakukan prediksi pendapatan dengan algoritma C4.5," *Go Infotech: J. Ilm. STMIK AUB*, vol. 30, no. 1, pp. 111–121, Jun. 2024, doi: 10.36309/goi.v30i1.266.

[13] R. Mukhaimin, R. Ester, and I. T. Hardono, "Optimalisasi business intelligence dengan metode CRISP-DM pada Departemen Service Solution PT Home Center Indonesia," *JRIIN: J. Ris. Inform. dan Inov.*, vol. 3, no. 5, pp. 1268–1272, 2025.

[14] D. Kurniawan and M. Yasir, "Optimization sentiment analysis using CRISP-DM and Naive Bayes methods implemented on social media," *Cyberspace: J. Pendidik. Teknol. Inf.*, vol. 6, no. 2, pp. 74–85, Oct. 2022, doi: 10.22373/cj.v6i2.12793.

[15] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic minority over-sampling technique," *J. Artif. Intell. Res.*, vol. 16, pp. 321–357, Jun. 2002, doi: 10.1613/jair.953.

[16] B. C. Perdana and E. A. Farida, "Komparasi algoritma Decision Tree dan Random Forest dalam prediksi kelayakan mahasiswa penerima beasiswa," *JUSINFO: J. Sist. Inf.*, vol. 1, no. 1, pp. 19–24, Nov. 2025.

[17] S. Budiman, A. Sunyoto, and A. Nasiri, "Analisa performa penggunaan feature selection untuk mendeteksi intrusion detection systems dengan algoritma Random Forest classifier," *Sistemasi: J. Sist. Inf.*, vol. 10, no. 3, pp. 754–760, Sep. 2021, doi: 10.32520/stmsi.v10i3.1550.

[18] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, Oct. 2001, doi: 10.1023/A:1010933404324.

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

[19] N. A. Amalia, I. T. Utami, and Y. Wilandari, "Analisis sentimen kebijakan penyelenggara sistem elektronik lingkup privat menggunakan penalized logistic regression dan Support Vector Machine," *J. Gaussian*, vol. 12, no. 4, pp. 560–569, 2023, doi: 10.14710/j.gauss.12.4.560-569.

[20] C. Cortes and V. Vapnik, "Support-vector networks," *Mach. Learn.*, vol. 20, no. 3, pp. 273–297, Sep. 1995, doi: 10.1007/BF00994018.

[21] R. Blanquero, E. Carrizosa, P. Ramírez-Cobo, and M. R. Sillero-Denamiel, "Variable selection for Naïve Bayes classification," *Comput. Oper. Res.*, vol. 135, Art. no. 105456, Nov. 2021, doi: 10.1016/j.cor.2021.105456.

[22] M. D. Muafa and L. Iswari, "Pengembangan aplikasi berbasis web dengan RShiny untuk data klasifikasi menggunakan metode Naive Bayes," *Automata*, vol. 3, no. 1, pp. 8–15, Jan. 2022.

[23] H. M. Nawawi, A. B. Hikmah, A. Mustopa, and G. Wijaya, "Model klasifikasi machine learning untuk prediksi ketepatan penempatan karir," *J. Saintekom*, vol. 14, no. 1, pp. 13–25, Mar. 2024, doi: 10.33020/saintekom.v14i1.512.