M A T H L I N E
JURNAL MATEMATIKA DAN PENDIDIKAN MATEMATIKA

ISSN 2502-5872 (Print)
ISSN 2622-3627 (Elektronik)

Volume 10 Number 1, February 2025, 221-235

CLASSIFICATION OF KIP-K SCHOLARSHIP USING LOGISTIC REGRESSION, CLASSIFICATION TREES, AND BOOSTING BASED ON DECISION SUPPORT SYSTEM

Uqwatul Alma Wizsa[1*], Alya Rahmi[2]
[1,3]Departement Statistics, UIN Sjech M. Djamil Djambek Bukittinggi, West Sumatera, Indonesia
*Correspondence: uqwatulalmawizsa@uinbukittinggi.ac.id

ABSTRACT

This study addresses the challenge of accurately identifying eligible awardees of the KIP-K scholarship at UIN Sjech M. Djamil Bukittinggi, where scholarship aid requests exceed the allocated funds. The research aims to develop an integrated classification and decision-making model to optimize the selection process. From the 2022 and 2023 scholarship applicant data obtained through AKAMA, preprocessing was conducted, resulting in a final dataset comprising 2,144 records. The dataset includes 14 explanatory variables influencing scholarship eligibility. The study compares three classification methods—logistic regression, classification tree, and boosting - using the 2022 data for training and testing. The SMOTE resampling technique was applied to address class imbalance. The novelty of this research lies in integrating classification analysis with a decision-making system based on the Simple Additive Weighting (SAW) method, enhancing the ranking of applicants based on criteria. The results indicate that logistic regression delivered the best performance in terms of accuracy, sensitivity, and AUC-ROC scores during testing, despite a slight decline in performance when applied to the 2023 dataset. Moreover, integrating logistic regression with SAW significantly improved decision-making precision. The application of logistic regression combined with SAW on the 2023 dataset resulted in a final accuracy of 0.5734 and a balanced accuracy of 0.5820. This integrated framework provides a data-driven, fair, and efficient approach to scholarship allocation. The study highlights the importance of combining predictive models with decision-making systems to ensure equitable and targeted distribution of financial aid to deserving students.

Keywords: KIP-K Scholarship, logistic regression, decision-making, SMOTE.

How to Cite: Wizsa, U. A., & Rahmi, A. (2025). Classification Of KIP-K Scholarship Using Logistic Regression, Classification Trees, and Boosting Based On Decision Support System. *Mathline: Jurnal Matematika dan Pendidikan Matematika*, *10*(1), 221-235. http://doi.org/10.31943/mathline.v10i1.837

PRELIMINARY

Kartu Indonesia Pintar-Kuliah (KIP-K) is a scholarship program initiated by the Indonesian government to enhance access to higher education for students from underprivileged families. Through this program, students are exempted from tuition fees and receive financial support for daily needs. The initiative plays a critical role in reducing educational inequality and ensuring equitable opportunities for all.

At UIN Sjech M. Djamil Djambek Bukittinggi, the selection process for KIP-K recipients has been implemented using a Decision Support System (DSS) based on the Simple Additive Weighting (SAW) method. This method ranks applicants based on predefined criteria such as academic performance, family income, and other socio-economic indicators. This aligns with research by Wizsa et al who examined DSS for selecting KIP-K scholarship awardees using the SAW and TOPSIS methods. This research compares the two methods most frequently used in DSS and it is found that the SAW method is the best method that can be applied in the selection of KIP Kuliah scholarship (Wizsa et al., 2022).

The ranking of scholarship candidates within DSS employs weight normalization on each decision-making criterion (Darlinda & Utamajaya, 2022). Thus, the decision-making process is solely based on the weights assigned to each criterion. This decision-making procedure could be improved if it begins with the classification process of the unified scholarship applicant data (Pendiagnosa et al., 2011). The data can be classified using modeling methods such as logistic regression, classification trees, and boosting (Wibowo & Djafar, 2023). This process divides observations into groups based on their characteristics. Applicants are initially classified as students deemed eligible and ineligible for the scholarship (Widianta et al., 2018). Subsequently, based on this grouping, a Decision Support System (DSS) using the SAW method is implemented to determine scholarship nominations based on the weight criteria set by scholarship administrators (Darlinda & Utamajaya, 2022). Additionally, relying on classification and decision support systems for scholarship selection is expected to assist administrators in screening applicants accurately (Natalis & Nataliani, 2022).

In this study, logistic regression, classification trees, and boosting were chosen for their complementary strengths. Logistic regression offers simplicity and interpretability for binary outcomes. Classification trees are capable of capturing complex, non-linear relationships, making them ideal for datasets with intricate patterns. Boosting, particularly Extreme Gradient Boosting (XGBoost), enhances prediction accuracy by iteratively minimizing errors, which is crucial for high-stakes decisions like scholarship selection (Chen & Carlos, 2016). Alternative methods such as Naïve Bayes and Random Forest were considered but excluded. Naïve Bayes, while efficient, assumes conditional independence among predictors, which is often unrealistic in real-world datasets. Random Forest, though robust, is computationally intensive and less interpretable, posing challenges in communicating results to stakeholders.

Rini Sovia used the K-Means algorithm and the SAW Decision Support System to predict achievement-based scholarship awardees. In the study, the researcher applied a single classification approach using K-Means and the SAW method for DSS SAW. This study produced a user-friendly application system that can be used in the selection process for achievement-based scholarships at schools, as well as for determining foundation scholarship awardees using the K-Means classification method (Sovia et al., 2020). Further, Sudarsono's study categorized scholarship candidates into four groups based on specified criteria (Sudarsono & Lestari, 2021). In 2020, Aah Sumiah compared the K-Nearest Neighbor and Naïve Bayes methods for determining scholarship awardees. This research found a high data accuracy rate, with 100% for the KNN method and 99.98% for Naïve Bayes. This study resulted in an information system that can be used by academic departments as a recommendation in the scholarship selection process (Sumiah & Mirantika, 2020). Additionally, research on the classification of Bidikmisi scholarship awardees across East Java employed oversampling, undersampling, SMOTE, SVM, and Random Forest. The researchers found that the application of random oversampling and SMOTE yielded nearly identical AUC values, which are suitable for handling imbalanced data cases (Qadrini et al., 2022).

In prior studies, no research was found comparing logistic regression, classification tree, and boosting methods with the SAW Decision Support System. The researcher considers it necessary to compare the classification process using multiple methods to determine which yields better predictions. Furthermore, combining the classification process with the SAW Decision Support System is conducted to align decision-making criteria established by scholarship administrators in determining eligible scholarship awardees.

METHODS

Data

The data used in this study comprises secondary data on applicants for the Kartu Indonesia Pintar-Kuliah (KIP-K) scholarship at UIN Sjech M. Djamil Djambek Bukittinggi for 2022-2023, obtained from the academic and student affairs department (AKAMA). The data includes one response variable and 14 explanatory variables. The total number of students who applied for the KIP-K scholarship in 2022 was 1,303, while in 2023, it was 842. After data preparation, such as cleaning out any missing data, a total of 2,144 scholarship applicants' data were included in the study.

Mechanism of Research

This research used a descriptive-quantitative method with machine-learning approach using R. The first step was data exploration to obtain a general overview of the data to be analyzed. Then, the data was split into two parts. First, the data from 2022 applicants who had been accepted or rejected as scholarship awardees. This data was used to develop the best classification model. The second set of data comprised 2023 applicants who had not yet received a decision on their scholarship status. This second set of data would then be used to predict scholarship acceptance or rejection. The 2022 data (first dataset) was further split, with 75% randomly assigned as training data and the remaining 25% as testing data. The training data was then oversampled to balance the data.

The balanced training data from the oversampling process was analyzed using various classification modeling methods, with model performance assessed through testing data. Next, the best classification model was used to predict the response variable (acceptance or rejection) for the 2023 applicant data. Scholarship applicants deemed eligible (accepted) in the classification results were ranked using the SAW Decision Support System (DSS) method. Finally, a confusion matrix was constructed to compare the model's predictions for 2023 scholarship awardees with the actual data.

Method

Classification modeling is a set of rules that can be used to determine or place an object into a specific class or group. These rules are derived from data of other objects whose classes are already known. In machine learning discussions, classification models are models or rules obtained through supervised learning, where the response variable is the class(Mandaku & Mandaku, 2010).

The rules for classifying new objects are derived from data containing many objects with known classes. Besides information about the classes, this data also includes variables that characterize the objects in the dataset. Classification modeling generally works to recognize distinguishing rules or functions between classes. The existence of these distinguishing rules or functions is useful in determining the class or category of new objects whose classes are unknown. In other words, a classification model can be useful in predicting the class of an observation whose class is not yet known (Sartono & Dharmawan, 2023).

In the classification field, there is a term called class imbalance, which refers to an unequal number of observations in each category, resulting in a majority class (a large amount of data) and a minority class (a small amount of data). While differences in the number of observations across categories are permissible, significant imbalances can impact estimation results. Although it may yield highly accurate predictions for the majority class, using such imbalanced data directly for predictions often leads to inaccurate or even completely failed predictions for the minority class (Lin et al., 2017).

Imbalanced data conditions must be addressed, one method being resampling. Resampling means replacing the unbalanced dataset with a new balanced dataset. One resampling procedure that can be used is oversampling. Applying oversampling can improve measurement results in classification analysis. One popular oversampling technique is the Synthetic Minority Over-Sampling Technique (SMOTE), which synthesizes new samples from the minority class to balance the dataset by resampling minority class samples (Siringoringo, 2018). This approach ensures that the model does not overfit the majority class while improving the classification accuracy of the minority class.

One of the most commonly used classification algorithms is logistic regression. This algorithm can separate the dataset into two parts, known as binary classification (Satyahadewi et al., 2023). Logistic regression models produce categorical and qualitative output (Primartha & Wahono, 2021). The logistic regression graph divides the dataset into two classes (class=1 and class=0) right in the middle, at y=0.5. The classes are determined based on probability predictions (p) where $p \geq 0{,}5$ assigns class=1 and $p < 0{,}5$ assigns class=0 (Purwa, 2019)(Speelman, 2014). Logistic regression provides a model called the logit model, which yields a probability value (Purwa, 2019). Logistic regression uses the logit function to transform linear values into probabilities. The basic equation of logistic regression is:

$$P(y = 1|X) = \frac{e^{\beta_0 + \beta_1 X_1 + \cdots + \beta_p X_p}}{1 + e^{\beta_0 + \beta_1 X_1 + \cdots + \beta_p X_p}} \quad (1)$$

The value $P(y = 1|X)$ is the probability that the dependent variable y equals 1 (event), β_0 is the constant, and $\beta_1, \beta_2, \dots, \beta_n$ are the regression coefficients associated with each independent variables $X_1, X_2, \dots, X_n$. The probability value ranges between zero and one, $0 < P(y = 1|X) < 1$.

The second method is a classification tree. Classification trees are a classification algorithm that can be used to make decisions based on observations. The algorithm's response can be a categorical variable. In its decision-making process, this algorithm divides the training data with known response variable information based on the homogeneity of responses by category, recursively. At the initial stage, data is divided into two parts by separating the first category from the second based on the dominant homogeneity. In the next stage, the result of this division is further divided into two parts in the same manner. This step is repeated until the partitioning is maximized, ensuring that the two categories are well separated within each partition (Sartika & Sensuse, 2017).

Classification trees can predict the class of an observation whose class is unknown. Class predictions are based on the characteristics of the observation according to the values of explanatory variables. Moreover, classification trees can identify class characteristics and differences between class characteristics. Finally, classification trees can reveal important variables for identifying the segments or classes of observations (Sartono & Dharmawan, 2023).

The criterion for determining whether partitioning is maximized can be derived from the entropy value, denoted as E(D), defined for binary classification with

$$E(D) = -p_0 \log_2 p_0 - p_1 \log_2 p_1 \quad (2)$$

that p_0 as the probability of class 0 and p_1 as the probability of class 1.

The last method is boosting. Boosting is a machine learning technique used to improve model prediction performance. The basic principle of boosting is to combine several weak learners into a single strong learner. The Boosting model is a type of classification tree model enhancement. In boosting models, predictions obtained from the classification tree model are evaluated by calculating the prediction error (Herni Yulianti et al., 2022). The prediction error is calculated by identifying observations with incorrect predictions compared to their true classifications. Each observation is then weighted, where misclassified observations receive higher weights than correctly classified ones (Noroozi et al., 2018).

Extreme Gradient Boosting (XGBoost) is a popular implementation of the boosting algorithm, developed for high efficiency and performance, especially in handling large and complex data. XGBoost operates by using the gradient boosting method, an advancement of traditional boosting. At each iteration, XGBoost builds a new model based on the gradient or derivative of the loss function to minimize prediction errors (Chen & Carlos, 2016).

This model improves predictions by adding a new decision tree that predicts errors or residuals from the previous model. Unlike other boosting algorithms, XGBoost adds a regularization mechanism to reduce overfitting, making it more reliable for generalizing new data (Natekin & Knoll, 2013).

After the classification process, scholarship candidates classified as passing in the classification results are then ranked using Simple Additive Weighting (SAW). The Simple Additive Weighting (SAW) method is one of the Decision Support System (DSS) methods. The SAW method normalizes the decision matrix to a comparable scale with all available alternatives (Piasecki & Roszkowska, 2019). The calculation stages of the SAW method begin with determining the alternatives, defining evaluation criteria, assessing the suitability of each alternative for each criterion, creating a normalized matrix, forming a normalized matrix, and finally ranking each alternative to achieve a decision result (Khasanah & Rofiah, 2019).

After classifying a specific dataset cluster, it is essential to evaluate the prediction performance of the classification model. To assess whether the model effectively classifies observational data, a comparison between predicted and actual classes can be conducted (Purwa, 2019). Five common criteria in classification modeling are accuracy, sensitivity, specificity, precision, and balanced accuracy. These criteria can be calculated using a cross-tabulation between actual and predicted classes.

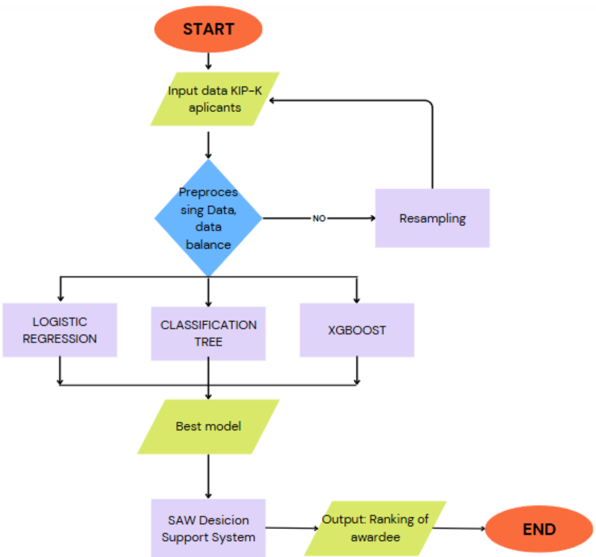


Figure 1. Flowchart of the analytical in research

RESULT AND DISCUSSION

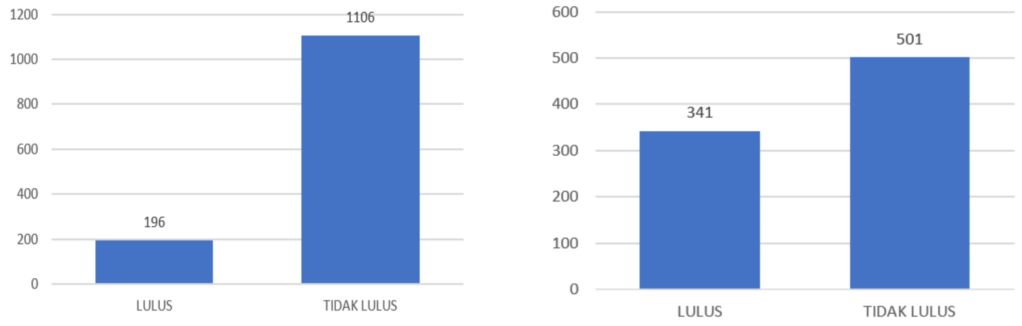
Data Exploration

This study used data from applicants for the KIP-K scholarship in 2022 and 2023. There were 1,302 applicants in 2022 and 839 in 2023. Each dataset included a single response variable indicating scholarship acceptance status (accepted or rejected). The explanatory variables used as criteria for student scholarship acceptance were drawn from 14 detailed variables explained in Table 1.

Table 1. Variables of data

No	Type	Name
1	Response variable	Kelulusan (Y)
2		Penghasilan Orang Tua (X1)
3		Pendidikan Ayah (X2)
4		Pendidikan Ibu (X3)
5		Pekerjaan Ayah (X4)
6		Pekerjaan Ibu (X5)
7		Status Orang Tua (X6)
8	Explanatory variables	Jumlah Tanggunggan (X7)
9		Raskin (X8)
10		KIP (X9)
11		Rapor (X10)
12		Tahfiz (X11)
13		Prestasi (X12)
14		Status Rumah (X13)
15		Daya Listrik Rumah (X14)

Figure 2 shows bar chart of KIP-K scholarship acceptance of 2022 and 2023. In 2022, where out of 1,302 applicants, only 196 were accepted, and 1,106 were rejected, with a 15% acceptance rate (minor category) and an 85% rejection rate (major category). In 2023, the acceptance count increased to about 341, making it less skewed compared to the 501 rejections.



(a) (b)

Figure 2. Bar Chart of KIP-K Scholarship 2022 (a) and 2023 (b) using Ms. Excel Result

For the classification of scholarship applicants, data from the last two years was used: 2022 data as training data and 2023 data for prediction. The 2022 data was used as training data to identify the best classification method. This dataset was initially split into training and testing data to analyze the characteristics of accepted and rejected applicants. This information was then used to run the classification algorithm and determine the best model for classifying scholarship awardees. The dataset was divided into 75% training and 25% testing, yielding 147 accepted and 830 rejected applicants in the training data.

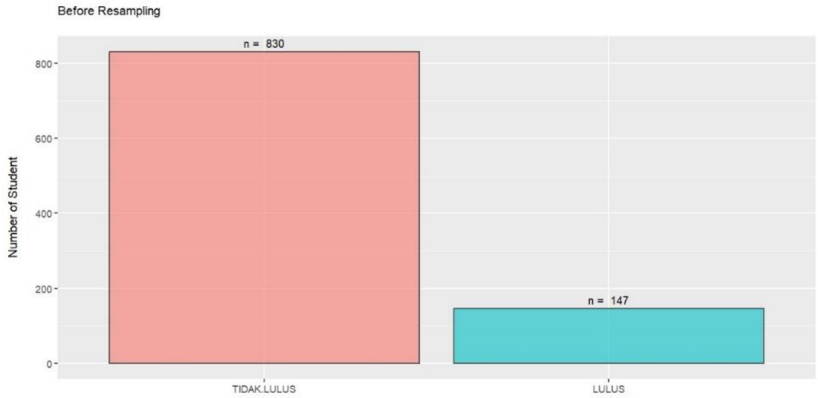


Figure 3. Data training bar chat before resampling using R

It was observed from Figure 3 that the 2022 dataset revealed significant class imbalance which posed challenges for training classification models, as imbalanced data often biases predictions toward the majority class. To address this, the Synthetic Minority Over-Sampling Technique (SMOTE) was applied to balance the dataset by generating synthetic samples for the minority class through interpolation between existing data points. Unlike undersampling, which risks losing valuable information from the majority class, or more advanced resampling techniques, SMOTE was chosen for its simplicity and proven effectiveness in educational datasets. Balancing the data using SMOTE improved model performance, particularly in predicting the minority class, as reflected in enhanced metrics like sensitivity and balanced accuracy. This step significantly mitigated prediction bias and ensured a more reliable decision-making process for KIP-K scholarship awardee selection. The imbalanced data was resampled using the SMOTE method, resulting in 588 accepted and 661 rejected cases. The data distribution is shown in Figure 4.

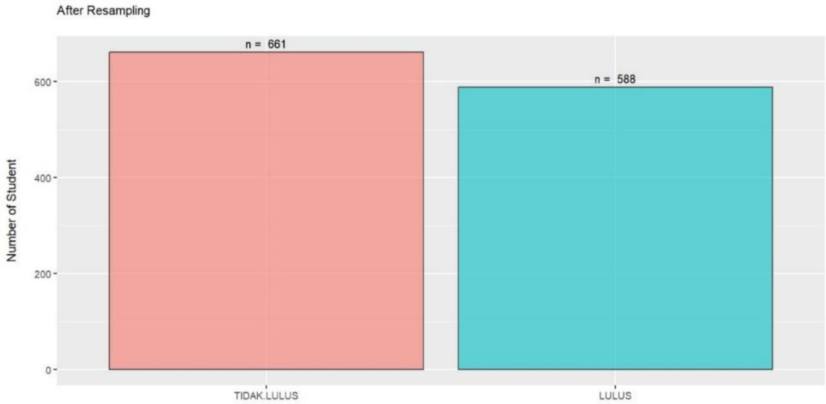


Figure 4. Data training bar chart after resampling using R

The resampled data was then analyzed using logistic regression, classification trees, and boosting algorithms with the XGBoost model. The performance of each model is detailed in Table 2, which shows sensitivity at 0.8571, specificity at 0.9203, accuracy at 0.9108, balanced accuracy at 0.8887, and AUC-ROC at 0.8147. Based on these values, logistic regression was identified as the best model.

Table 2. The best model criteria

Criteria	Model		
	Logistic regression	Classification tree (CART)	XGboost
Sensitivity	**0.8571**	0.8367	0.8163
Specificity	**0.9203**	0.8514	0.9130
Accuracy	**0.9108**	0.8492	0.8985
Balanced Accuracy	**0.8887**	0.8441	0.8647
AUC-ROC	**0.8147**	0.7335	0.7952

Once logistic regression was identified as the best model, it was used to predict the scholarship acceptance for 2023 applicants. The logistic regression model metrics on the 2023 data are presented in Table 3. There was a decrease in sensitivity, specificity, accuracy, balanced accuracy, and AUC-ROC values on the 2023 applicant data, with sensitivity at 0.2405, specificity at 0.8603, accuracy at 0.6093, balanced accuracy at 0.5504, and AUC-ROC at 0.5821.

Table 3. The model criteria for the 2023 data

Criteria	Logistic Regression
Sensitivity	0.2405
Specificity	0.8603
Accuracy	0.6093
Balanced Accuracy	0.5504
AUC-ROC	0.5821

The significant drop in logistic regression performance on the 2023 dataset, particularly the decline in sensitivity from 0.8571 to 0.2405, suggests potential overfitting to the 2022 data or differences in data distribution between the two years. Overfitting occurs when the model captures noise or specific patterns unique to the training data, reducing its ability to generalize to new data. Alternatively, the 2023 dataset may have introduced new applicant characteristics or shifted class distributions that were not represented in the 2022 dataset. To mitigate this issue, future work should incorporate techniques such as cross-validation, regularization, or domain adaptation methods to improve the model's robustness. Additionally, monitoring changes in data patterns across years and retraining models with updated datasets can enhance generalizability and maintain consistent performance over time.

The predicted scholarship acceptance results using logistic regression were compared with the actual 2023 data, as shown in Table 4. There were 82 applicants correctly predicted as accepted, and 259 correctly predicted as rejected. There were also 70 predicted as accepted who were actually rejected, and 431 predicted as rejected who were actually rejected.

Tabel 4. Confusion matrix of prediction for 2023 data using logistic regression

Prediction	Actual	
	Accepted	Rejected
Accepted	82	70
Rejected	259	431

Using the logistic regression model, an analysis of important explanatory variables and their categories for determining scholarship eligibility was conducted. The top five important variables were: "raskin" (poverty assistance program) with the "not available" category, "tahfiz" (Quran memorization) with the "not available" category, "homeownership" with the "self-owned, paid off" category, "number of dependents" with the "more than five" category, and "grades" with the 80-90 category. These variables with their categories can classify the scholarship acceptance status.

Table 5. Important explanatory variables

Variabel Penjelas	Overall
Raskin (tidak ada)	100.00
Tahfiz (tidak ada)	73.49
Status rumah (milik sendiri lunas)	69.60
Jumlah Tanggungan (>5)	47.14
Rapor (80-89)	41.30
Status rumah (milik saudara)	39.34
Prestasi (tidak memiliki)	36.27
Status rumah (milik sendiri belum lunas)	36.18
Rapor (>=90)	34.64
Jumlah tanggungan (4-5)	30.78
Pendidikan ayah (SMA/sederajat)	24.99
Prestasi (Kabupaten/kota)	24.32
Pekerjaan ayah (wiraswasta)	24.04
Pendidikan ayah (SMP/sederajat)	19.48
Pendidikan ayah (SD/TT SD)	17.04
Pekerjaan Ibu (meninggal/tidak ada pekerjaan)	15.91
Pekerjaan Ayah (tidak bekerja)	15.24
Prestasi (provinsi/nasional)	14.45
Pekerjaan ayah (nelayan)	13.42
Pekerjaan ibu (wiraswasta)	13.33

Finally, the classification results using logistic regression were applied to the Simple Additive Weighting (SAW) Decision Support System to rank the top 504 scholarship applicants. Table 6 presents the 2023 scholarship applicant predictions using this combined model. The combined model showed improved predictions compared to logistic regression alone, with 183 applicants predicted and actually accepted, 57 predicted rejected but actually accepted, 158 predicted accepted but actually rejected, and 106 predicted and actually rejected.

Table 6. Confusion matrix of prediction for 2023 data used logistic regression-SAW

Prediction	Actual	
	Accepted	Rejected
Accepted	183	158
Rejected	57	106

In general, based on the evaluation metrics, there is an observed increase in the balanced accuracy of the combined model, reaching 0.5820, compared to the regression-only model, which achieved 0.5504. This aligns with Bishop's assertion (Bishop, 2006) that machine learning, particularly classification, can yield better and more efficient results in data cluster separation.

Table 7. The model criteria using model logistic regression-SAW

Criteria	Mix Model
Sensitivity	0.7625
Specificity	0.4015
Accuracy	0.5734
Balanced Accuracy	0.5820

Although this study obtained model goodness values from the combination of logistic regression and the SAW method, with an accuracy of 0.5734 and a balanced accuracy of 0.5820, these values are relatively low. Therefore, further research is needed to improve accuracy so that the implemented model can be more effective and precise in determining scholarship recipients.

According to Aikia et al. (2021), achieving higher accuracy in classification models is crucial, especially in sensitive decision-making systems like scholarship selection(Aikia, R., Roy, S., & Bhattacharya, 2021). Models with lower accuracy can lead to misclassification, affecting fairness and efficiency. Similarly, Sharma and Kumar (2020) emphasize that accuracy values above 0.70 are generally preferred in predictive modeling to ensure reliability and trustworthiness in practical applications. Thus, improving the model's accuracy is essential for better performance and decision support (Sharma, R., & Kumar, 2020).

CONCLUSION

In conclusion, the analysis found that the logistic regression model is the most accurate and reliable method for classifying eligible KIP-K scholarship awardees, outperforming classification trees and boosting techniques in terms of sensitivity, specificity, accuracy, balanced accuracy, and AUC-ROC values. The logistic regression model's robust performance highlights its effectiveness in handling imbalanced data and predicting eligibility based on key criteria, such as family income, educational background of parents, and student achievements. This model's ability to differentiate applicants with high accuracy ensures that the selection process can more effectively identify students most in need of financial aid.

Furthermore, the integration of the Simple Additive Weighting (SAW) decision support system with the classification model significantly enhances the scholarship selection process. The SAW method enables precise ranking of applicants by weighing critical criteria, which, when combined with the logistic regression classification results, offers a comprehensive approach for the scholarship committee. This integrated framework streamlines decision-making, ensuring a fair and efficient selection of KIP-K awardees, and potentially setting a benchmark for similar scholarship programs aiming to improve access to higher education for underprivileged students.

ACKNOWLEDGMENT

We would like to extend our sincere gratitude to the leadership of UIN Sjech M. Djamil Djambek Bukittinggi for their invaluable support, and to the Litapdimas BOPTN program for providing the research grant funding. This study would not have been possible without the generous assistance and resources provided, enabling us to carry out this research and contribute to the advancement of knowledge in the field of applied statistics.

REFERENCES

Aikia, R., Roy, S., & Bhattacharya, A. (2021). Improving Classification Accuracy in Decision-Making Systems: A Review. *International Journal of Data Science and Analysis*, *7*(4), 123–135.

Bishop, C. M. (2006). *Pattern Recognition and Machine Learning* (1st ed.). Springer.

Chen, & Carlos, T. and G. (2016). *XGBoost: A Scalable Tree Boosting System*. https://doi.org/10.1145/2939672.2939785

Darlinda, D., & Utamajaya, J. N. (2022). Sistem Pendukung Keputusan Penerima Beasiswa Program Indonesia Pintar Menggunakan Metode Algoritma K-Means

Clustering. *JURIKOM (Jurnal Riset Komputer)*, *9*(2), 167. https://doi.org/10.30865/jurikom.v9i2.3971

Herni Yulianti, S. E., Oni Soesanto, & Yuana Sukmawaty. (2022). Penerapan Metode Extreme Gradient Boosting (XGBOOST) pada Klasifikasi Nasabah Kartu Kredit. *Journal of Mathematics: Theory and Applications*, *4*(1), 21–26. https://doi.org/10.31605/jomta.v4i1.1792

Khasanah, F. N., & Rofiah, S. (2019). Sistem Seleksi Penerimaan Beasiswa Menggunakan Metode Pendukung Keputusan Simple Additive Weighting. *Seminar Nasional APTIKOM (SEMNASTIK)*.

Lin, W. C., Tsai, C. F., Hu, Y. H., & Jhang, J. S. (2017). Clustering-based undersampling in class-imbalanced data. *Information Sciences*, *409–410*, 17–26. https://doi.org/10.1016/j.ins.2017.05.008

Mandaku, F., & Mandaku, H. (2010). Pemodelan Terhadap Kelulusan Siswa Masuk Kelas Akselerasi Menggunakan Analisis Regresi Logistik Dan Multivariate Adaptive Regression Spline (Mars). *Arika*, *04*(2), 159–166. https://ojs3.unpatti.ac.id/index.php/arika/article/view/473

Natalis, A., & Nataliani, Y. (2022). Pemanfaatan k-Means Clustering dan Analytic Hierarchy Process terhadap Penilaian Prestasi Kerja Pegawai. *Jurnal Teknik Informatika Dan Sistem Informasi*, *8*(1), 88–99. https://doi.org/10.28932/jutisi.v8i1.4243

Natekin, A., & Knoll, A. (2013). Gradient boosting machines, a tutorial. *Frontiers in Neurorobotics*, *7*(DEC). https://doi.org/10.3389/fnbot.2013.00021

Noroozi, M., Vinjimoor, A., Favaro, P., & Pirsiavash, H. (2018). Boosting Self-Supervised Learning via Knowledge Transfer. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 9359–9367. https://doi.org/10.1109/CVPR.2018.00975

Pendiagnosa, A., Warna, K., Pemrograman, M., Delphi, B., & Eniyati, S. (2011). Perancangan Sistem Pendukung Pengambilan Keputusan untuk Penerimaan Beasiswa dengan Metode SAW (Simple Additive Weighting). *Jurnal Teknologi Informasi DINAMIK*, *16*(2), 171–176. https://www.unisbank.ac.id/ojs/index.php/fti1/article/view/364

Piasecki, K., & Roszkowska, E. (2019). *SS symmetry Simple Additive Weighting Method Equipped with*.

Primartha, R., & Wahono, R. S. (2021). *Algoritma Machine Learning*. Informatika Bandung.

Purwa, T. (2019). Perbandingan Metode Regresi Logistik dan Random Forest untuk Klasifikasi Data Imbalanced (Studi Kasus: Klasifikasi Rumah Tangga Miskin di Kabupaten Karangasem, Bali Tahun 2017). *Jurnal Matematika, Statistika Dan Komputasi*, *16*(1), 58. https://doi.org/10.20956/jmsk.v16i1.6494

Qadrini, L., Hikmah, H., & Megasari, M. (2022). Oversampling, Undersampling, Smote SVM dan Random Forest pada Klasifikasi Penerima Bidikmisi Sejawa Timur Tahun 2017. *Journal of Computer System and Informatics (JoSYC)*, *3*(4), 386–391. https://doi.org/10.47065/josyc.v3i4.2154

Sartika, D., & Sensuse, D. I. (2017). Perbandingan Algoritma Klasifikasi Naive Bayes, Nearest Neighbour, dan Decision Tree pada Studi Kasus Pengambilan Keputusan Pemilihan Pola Pakaian. *Jatisi*, *1*(2), 151–161. https://jurnal.mdp.ac.id/index.php/jatisi/article/view/78

Sartono, B., & Dharmawan, H. (2023). *Pemodelan Prediktif Berbasis Pohon Klasifikasi*.

IPB Press.

Satyahadewi, N., Tamtama, R., Perdana, H., & Huriyah, S. K. (2023). Binary Logistics Regression To Predict The Opportunity Of SNMPTN Graduation In Statistics Study Program Of Tanjungpura University. *Mathline : Jurnal Matematika Dan Pendidikan Matematika*, *8*(1), 17–28. https://doi.org/10.31943/mathline.v8i1.269

Sharma, R., & Kumar, P. (2020). Evaluating Predictive Models: Accuracy, Precision, and Beyond. *Journal of Machine Learning Applications*, *5*(2), 45–59. https://ieeexplore.ieee.org/document/9311735

Siringoringo, R. (2018). Klasifikasi Data Tidak Seimbang Menggunakan Algoritma SMOTE dan k-Nearest Neighbor. *Jurnal ISD*, *3*(1), 44–49. https://ejournal-medan.uph.edu/isd/article/view/177

Sovia, R., Mandala, E. P. W., & Mardhiah, S. (2020). Algoritma K-Means dalam Pemilihan Siswa Berprestasi dan Metode SAW untuk Prediksi Penerima Beasiswa Berprestasi. *Jurnal Edukasi Dan Penelitian Informatika (JEPIN)*, *6*(2), 181. https://doi.org/10.26418/jp.v6i2.37759

Speelman, D. (2014). Logistic regression: A confirmatory technique for comparisons in corpus linguistics. *Corpus Methods for Semantics: Quantitative Studies in Polysemy and Synonymy*, 487–533. https://doi.org/10.1075/hcp.43.18spe

Sudarsono, B. G., & Lestari, S. P. (2021). Clustering Penerima Beasiswa Yayasan Untuk Mahasiswa Menggunakan Metode K-Means. *Jurnal Media Informatika Budidarma*, *5*(1), 258. https://doi.org/10.30865/mib.v5i1.2670

Sumiah, A., & Mirantika, N. (2020). Perbandingan Metode K-Nearest Neighbor dan Naive Bayes untuk Rekomendasi Penentuan Mahasiswa Penerima Beasiswa pada Universitas Kuningan. *Buffer Informatika*, *6*(1), 1–10. https://journal.uniku.ac.id/index.php/buffer/article/view/2907

Wibowo, M., & Djafar, M. R. F. (2023). Perbandingan Metode Klasifikasi Untuk Deteksi Stress Pada Mahasiswa di Perguruan Tinggi. *Jurnal Media Informatika Budidarma*, *7*(1), 153–159. https://doi.org/10.30865/mib.v7i1.5182

Widianta, M. M. D., Rizaldi, T., Setyohadi, D. P. S., & Riskiawan, H. Y. (2018). Comparison of Multi-Criteria Decision Support Methods (AHP, TOPSIS, SAW & PROMENTHEE) for Employee Placement. *Journal of Physics: Conference Series*, *953*(1). https://doi.org/10.1088/1742-6596/953/1/012116

Wizsa, U. A., Yuspita, Y. E., & Rahayu, W. D. (2022). Decision-Making System for KIP IAIN Bukittinggi Scholarship Recipients Using the SAW and TOPSIS Methods. *Knowbase : International Journal of Knowledge in Database*, *2*(1), 85. https://doi.org/10.30983/ijokid.v2i1.5188