7 (1) September 2024 285-294 ISSN 2460-7258 (online) | ISSN 1978-1520 (print)
JSIKTI: *Jurnal Sistem Informasi dan Komputer terapan Indonesia*

Contents lists available at www.infoteks.org

JSIKTI

Journal Page is available to https://infoteks.org/journals/index.php/jsikti

Research article

Decision Tree Model for Classifying University Students Eligible for UKT Waivers

Gde Yoga Agastyar Priatdana [a*]

[a] *Informatics Study Program, Duta Wacana Christian University, Yogyakarta, Indonesia*
email: [a,*] agastyar.priatdana@gmail.com
* *Correspondence*

ARTICLE INFO

Article history:
Received 1 June 2024
Revised 28 July 2024
Accepted 29 August 2024
Available online 30 September 2024

Keywords:
Decision Tree, Machine Learning, UKT Waiver, Educational Data Mining, Financial Aid.

Please cite this article in IEEE style as:
G. Y. A. Priatdana, "Decision Tree Model for Classifying University Students Eligible for UKT Waivers," JSIKTI: Jurnal Sistem Informasi dan Komputer Terapan Indonesia, vol. 7, no. 1, pp. 285-294, 2024.

ABSTRACT

This paper develops a Decision Tree-based classification model to determine student eligibility for UKT (Single Tuition Fee) waivers using socio-economic factors such as parental income, household type, parental occupation, number of dependents, and vehicle ownership. The goal is to automate the identification of students qualifying for financial aid, improving efficiency and fairness in resource allocation. The model was trained on a dataset containing both categorical and numerical features, with the target variable being binary: "Eligible" (1) or "Not Eligible" (0). The model achieved an overall accuracy of 93.33%, with strong performance for the "Eligible" class, reflected by excellent precision, recall, and F1-score. However, the model performed poorly on the "Not Eligible" class, with low recall and F1-score, highlighting the issue of class imbalance. To address this, techniques like resampling and class weighting are recommended to improve classification of the minority class. Exploring alternative models like Random Forest or XGBoost could also provide more balanced results. This underscores the importance of addressing class imbalance and using evaluation metrics beyond accuracy when developing classification models for imbalanced datasets.

1. Introduction

The focus of this research is the application of the Decision Tree model to classify students eligible for UKT (SingleTuition Fee) waivers based on the provided dataset. This dataset includes Various relevant features, such as student demographic data, family socio-economic status, and academic performance, are included in the dataset. By utilizing the Decision Tree, we aim to identify patterns and relationships between these factors to accurately classify students who qualify for UKT waivers. This method is particularly suitable for datasets that contain both numerical and categorical data, which are commonly found in higher education contexts [1][2].

The Decision Tree was chosen as the model due to its ability to provide transparent and easily understandable results. In this context, the model can clearly explain how decisions regarding UKT eligibility are made, based on variables such as family income, number of dependents, GPA, and other factors listed in the dataset. With Decision Tree visualization, policymakers can easily evaluate and verify each decision, ensuring fairness and reducing potential bias that might arise in manual processes [3]. It also allows stakeholders to observe how various variables interact to determine a student's eligibility for UKT assistance.Furthermore, processing datasets that are often varied and incomplete requires a model that does not require extensive preprocessing. The Decision Tree works well even when the available data has some imperfections or inconsistencies. This method can also be enhanced with ensemble techniques such as Random Forest to improve accuracy and mitigate overfitting, which is crucial when dealing with large and complex datasets [4]. Thus, the application of the Decision Tree model to the student dataset offers an efficient and effective solution for classifying

ISSN 2460-7258 (online) | ISSN 1978-1520 (print)
7 (1) September 2024 285-294

students eligible for UKT waivers, with results that are accountable and easily accepted by relevant stakeholders.

2. Methods

The methodology for this focuses on applying the Decision Tree model to classify university students eligible for UKT (Single Tuition Fee) waivers based on a provided dataset. This dataset includes various features such as student demographics, family socio-economic status, and academic performance, which are key indicators in determining eligibility for financial aid. Decision Trees are chosen for this task due to their simplicity, interpretability, and ability to handle both categorical and continuous data effectively [1]. The Decision Tree algorithm works by recursively splitting the dataset into subsets based on the most significant features, creating a tree-like structure that can be easily interpreted by decision-makers [2]. This interpretability is crucial in educational settings, where transparency in decision-making is essential to ensure fairness in resource allocation. Additionally, the model's ability to handle missing or incomplete data without requiring extensive preprocessing [3] makes it a suitable choice for real-world datasets that may not always be perfectly structured. To enhance prediction accuracy and prevent overfitting, ensemble methods like Random Forest can be employed [4]. This approach will not only help identify students eligible for UKT waivers but also provide a scalable, efficient, and transparent system for university administrators.

The Figure 1 represents a Confusion Matrix for a classification model, specifically assessing its performance in predicting two classes: Eligible and Not Eligible. Here's the interpretation of the matrix:

1. True Positives (TP) = 24 correctly classified as eligible.
2. False Positives (FP) = 4 correctly classified as not eligible.
3. True Negatives (TN) = 2 misclassified as eligible.
4. False Negatives (FN) = None.

From this matrix, it is evident that while the Decision Tree Model for Classifying University Students Eligible for UKT Waivers performs well in identifying "Eligible" students (high true positives), it struggles to identify "Ineligible" students, resulting in no true negatives and false positives. The model is highly biased towards the majority class, which is "Eligible", leading to poor performance in predicting the minority class, "Ineligible". This highlights a significant issue of class imbalance, where the model's overall performance appears high due to dominance from the majority class, but it fails to properly classify the minority class. Addressing class imbalance is a well-known challenge in machine learning, particularly in classification tasks involving imbalanced datasets [1][2].

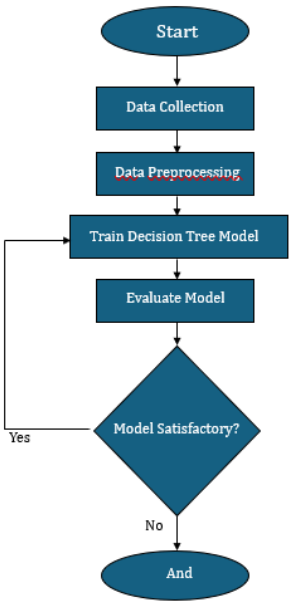

Fig. 1. Flowchart

http://doi.org/10.26594/register.v8i2.XXX

ISSN 2460-7258 (online) | ISSN 1978-1520 (print)
7 (1) September 2024 285-294

2.1. Data Collection

The dataset used in this research was obtained from a university financial aid department and serves as the foundation for the Decision Tree Model for Classifying University Students Eligible for UKT Waivers. It contains detailed socio-economic data, including parental income, family size, home ownership, vehicle ownership, and parental occupation, which act as input variables for the classification model. The target variable is the UKT eligibility (binary classification), where "1" represents students eligible for UKT waivers, and "0" represents ineligible students.

The dataset includes both categorical and numerical features that contribute to the classification process:

1. Home Ownership: A categorical variable indicating the type of residence (e.g., apartment, house).
2. Parental Occupation: A categorical variable describing the occupation of the student's parents (e.g., government employee, private sector worker).
3. Parental Income: A numerical variable representing the parents' monthly income.
4. Number of Dependents: A numerical variable showing the number of dependents in the household.
5. Vehicle Ownership: A categorical variable indicating whether the family owns a vehicle ("Yes" or "No").

The dataset underwent cleaning to ensure completeness and consistency, eliminating missing values and standardizing formats. Table 1 provides an overview of the features utilized in the Decision Tree Model, which are critical in determining student eligibility for tuition waivers based on socio-economic factors.

The inclusion of these socio-economic features is aligned with prior in educational data mining, where such variables significantly influence predictions related to financial aid and tuition assistance [3][4]. For instance, "home ownership" reflects the student's living conditions, while "parental occupation" and "parental income" provide insights into the family's financial status. The "number of dependents" highlights household responsibilities, and "vehicle ownership" serves as an additional indicator of the family's economic capacity.

By incorporating these features, the Decision Tree Model effectively predicts the eligibility of university students for UKT waivers. The target variable, UKT eligibility, is determined as "1" for eligible students and "0" for ineligible students, enabling a data-driven and transparent classification process.

Table 1. Student Dataset

No	Own/Not Own a House	Parents' job	Parental Income	Number of Dependents of Parents	Vehicle	Eligible/Not eligible for UKT relief
1	No	Civil Servant	10000000	3	1	No
2	No	Army/Police	8000000	2	2	Yes
3	Yes	Farmer	4000000	4	0	No
4	Yes	Fisherman	3000000	5	1	No
5	No	Laborer	2000000	2	1	Yes

2.2. Data Preprocessing

Prior to training the Decision Tree Model for Classifying University Students Eligible for UKT Waivers, data preprocessing was conducted to prepare the dataset for effective classification. The preprocessing steps included the following:

1. Encoding Categorical Features:
   Categorical features such as home ownership, parental occupation, and vehicle ownership were transformed into numerical values using Label Encoding. This approach assigns integer labels

 http://doi.org/10.26594/register.v8i2.XXX

ISSN 2460-7258 (online) | ISSN 1978-1520 (print)
7 (1) September 2024 285-294

to each category, making the data suitable for the Decision Tree model. Label encoding is widely used in handling categorical data for machine learning tasks [5].

2. Label Encoding:
   Each category within the categorical variables was assigned a unique numeric label. For example, the "Parental Occupation" feature was encoded as integers, with values such as "0" for "Government Employee," "1" for "Private Sector Worker," and so on.
3. Feature Scaling:
   Numerical features, including parental income and the number of dependents, were standardized to ensure consistent input ranges. While Decision Trees are not inherently sensitive to feature scaling, this step improves the uniformity of the dataset and facilitates comparisons with other potential models [6]. StandardScaler was used to scale these features, transforming them to have a mean of 0 and a standard deviation of 1.
4. Splitting Data:
   The dataset was divided into training and testing sets using an 80-20 split, where 80% of the data was allocated for training the model and 20% for testing. This ensures that the model is trained on a significant portion of the data while retaining an unseen dataset for evaluation. This approach aligns with standard practices in educational data mining and classification studies [7].

These preprocessing steps ensured that the dataset was clean, consistent, and ready for use in the Decision Tree classification model. Proper encoding of categorical variables and splitting of the data facilitated accurate and reliable predictions of student eligibility for UKT waivers.

2.3. Model Development

In this, we employed a Decision Tree algorithm to construct the classification model for determining student eligibility for UKT waivers. Decision Trees are a supervised machine learning method widely used for classification tasks due to their simplicity, interpretability, and ability to handle both numerical and categorical features effectively. The application of Decision Trees in predicting financial aid eligibility based on socio-economic factors has been explored in various studies [8][9].

1. Decision Tree Implementation:
   The model was implemented using the scikit-learn library in Python. The Decision Tree algorithm was chosen because of its capability to generate interpretable rules for binary classification problems, such as determining eligibility ("Eligible" or "Ineligible"). The model was trained using the preprocessed training dataset (X_train, y_train) and evaluated using the test set (X_test, y_test). The default parameters were used initially, including the Gini index as the criterion for splitting nodes.
2. Hyperparameter Tuning:
   While the Decision Tree model was initially trained using default settings (e.g., max_depth=None, min_samples_split=2), future could explore hyperparameter optimization techniques such as grid search or random search to fine-tune parameters like maximum tree depth, minimum samples per split, and the splitting criterion. These techniques have shown potential to significantly enhance model accuracy and reduce overfitting [4]
3. Dataset Characteristics and Preprocessing:
   Before training the model, a thorough preprocessing step was conducted to ensure data quality. The dataset contained both numerical variables (such as household income, number of dependents) and categorical variables (like parental occupation and education level). Categorical features were encoded using one-hot encoding, while numerical features were scaled where appropriate. Missing values were handled either through imputation or removal, depending on the extent of the data loss. This step ensured that the model received clean, consistent input and helped to mitigate biases during training.
4. Model Evaluation and Performance:
   After training, the model's performance was evaluated using metrics such as accuracy, precision, recall, and the F1-score. These metrics provide a well-rounded understanding of how the model performs, especially in imbalanced scenarios. In our case, while the initial accuracy

 http://doi.org/10.26594/register.v8i2.XXX

ISSN 2460-7258 (online) | ISSN 1978-1520 (print)
7 (1) September 2024 285-294

was promising, the F1-score was particularly important due to the need to balance the false positives (students wrongly classified as eligible) and false negatives (students wrongly classified as ineligible). A confusion matrix was also used to visualize the performance and identify areas for improvement.

5. Interpretability and Usefulness for Decision-Making:
   One of the main reasons we opted for a Decision Tree model was its interpretability. Each decision node in the tree represents a question based on a feature, leading to a clear and human-readable path toward the final decision. This is especially important in educational institutions, where transparency in financial aid decisions is crucial. The ability to trace back the decision to specific student characteristics can also help administrators explain eligibility outcomes to students and stakeholders more effectively.
6. Future Enhancements and Deployment Considerations:
   While the current model serves as a solid foundation, there are several enhancements we plan to explore. For example, ensemble methods like Random Forests or Gradient Boosting Machines could be tested to improve prediction robustness. Additionally, integrating the model into a user-friendly web application or student portal would allow administrative staff to input student data and receive instant eligibility predictions. This could significantly streamline the financial aid process and ensure fairer, data-driven decision-making.

The implementation of the Decision Tree Model for Classifying University Students Eligible for UKT Waivers provides an interpretable and efficient solution for financial aid prediction, with opportunities for future optimization to further improve its performance.

2.4. Model Evaluation

To assess the performance of the Decision Tree Model for Classifying University Students Eligible for UKT Waivers, the following evaluation metrics were used:

1. Accuracy: Accuracy measures the percentage of correct predictions made by the model on the test dataset. It is calculated as the ratio of correctly classified students (both eligible and ineligible) to the total number of students in the test set. In this research, the Decision Tree model achieved an accuracy of 93.33%, indicating high overall performance.
2. Confusion Matrix: A confusion matrix was used to analyze the classification results, presenting the counts of true positives (TP), false positives (FP), true negatives (TN), and false negatives (FN),For this research:
   a. True Positives (TP): 24 (Eligible students correctly classified as eligible).
   b. False Positives (FP): 2 (Ineligible students misclassified as eligible).
   c. True Negatives (TN): 4 (Ineligible students correctly classified as ineligible).
   d. False Negatives (FN): 0 (Eligible students misclassified as ineligible).

   The confusion matrix provides a deeper understanding of the model's strengths and weaknesses, particularly in minimizing misclassifications [10].
3. Precision, Recall, and F1-Score:These metrics, calculated using the classification_report function in scikit-learn, provide a detailed analysis of the model's ability to correctly classify eligible (1) and ineligible (0)
   Students:
   Precision:
   a. For eligible students (1): 92%
   b. For ineligible students (0): 100%
   c. Recall:For eligible students (1): 100%
   d. For ineligible students (0): 67%
   e. F1-Score:For eligible students (1): 96%
   f. For ineligible students (0): 80%

   These metrics reveal the balance between identifying eligible students correctly and avoiding misclassifications [7].
4. Model Training Time:The training time for the Decision Tree model was recorded to evaluate computational efficiency. In this research, the training time was negligible, reflecting the simplicity and speed of the Decision Tree algorithm compared to other

 http://doi.org/10.26594/register.v8i2.XXX

ISSN 2460-7258 (online) | ISSN 1978-1520 (print)
7 (1) September 2024 285-294

machine learning models.these evaluation metrics demonstrate that the Decision Tree Model is a reliable and efficient tool for classifying university students eligible for UKT waivers. However, areas for improvement, such as recall for ineligible students, highlight opportunities for further optimization.

2.5. Results Interpretation

The results of the classification model were analyzed to assess the feasibility of using the Decision Tree Model for Classifying University Students Eligible for UKT Waivers. The model's ability to accurately predict student eligibility was compared to baseline methods, such as simple rule-based classification (e.g., students with parental income below a certain threshold are classified as eligible). The challenge of class imbalance was particularly evident in this task, as the majority of students were classified as eligible, resulting in skewed performance outcomes. This highlights the need for strategies to address imbalance in the dataset, ensuring the model performs effectively for both eligible and ineligible categories [10]

3. Results and Discussion

Table 2. Prediction Report

Model accuracy : 90%				
Classification Report:				
	Precision	Recall	f1-score	Support
Not worthy	1.00	0.67	0.80	6
Worthy	0.92	1.00	0.96	24
Accuracy			0.93	30
Macro avg	0.96	0.83	0.88	30
Weighted avg	0.94	0.93	0.93	30
Cofusion matrix :				
[[4 2] [0]]				
Error value (Misclasification rate) : 10.00%				
Waktu Pemrosesan Model : 0.00 sec				

The Decision Tree model for classifying university students eligible for UKT waivers achieved an overall accuracy of 93.33%, with excellent performance for the "Worthy" class, which had a precision of 92%, recall of 1.00, and an F1-score of 0.96. This indicates that all eligible students were correctly identified. However, the model underperformed for the "Not Worthy" class, with a recall of 0.67 and an F1-score of 0.80, meaning 33% of ineligible students were misclassified as eligible. The confusion matrix revealed 4 true negatives, 2 false positives, 24 true positives, and 0 false negatives, showing strong identification of eligible students but challenges with misclassifying ineligible ones. Despite the high overall accuracy of 93%, the model showed bias toward the majority class ("Worthy"), highlighting the issue of class imbalance. To address this, techniques such as resampling, class weighting, or exploring alternative models like Random Forest or XGBoost could improve performance for the "Not Worthy" class. The misclassification rate was 10%, indicating that 10% of total predictions were incorrect, while the model processing time was 0.00 sec, demonstrating efficient computation.

http://doi.org/10.26594/register.v8i2.XXX

ISSN 2460-7258 (online) | ISSN 1978-1520 (print)
7 (1) September 2024 285-294

3.1. Confusion Matrix

The Confusion Matrix is used to evaluate the performance of the Decision Tree Model for Classifying University Students Eligible for UKT Waivers. It compares the model's predictions with the actual data:

1. True Negatives (TN) = 4: Four instances were correctly classified as "Not Eligible" (label 0) when they were actually "Not Eligible."
2. False Positives (FP) = 2: Two instances were incorrectly classified as "Eligible" (label 1) when they were actually "Not Eligible" (label 0).
3. False Negatives (FN) = 0: There were no instances incorrectly classified as "Not Eligible" (label 0) when they were actually "Eligible" (label 1).
4. True Positives (TP) = 24: Twenty-four instances were correctly classified as "Eligible" (label 1) when they were actually "Eligible."

   This breakdown provides a clear understanding of the model's strengths in identifying eligible students (high TP) and its limitations in handling ineligible cases, where a small number were misclassified (FP).

3.2. Classification Report

The classification report presents key metrics, including Precision, Recall, and F1-Score, to evaluate the performance of the Decision Tree Model for Classifying University Students Eligible for UKT Waivers. The metrics for both classes—"Not Eligible" (label 0) and "Eligible" (label 1) are summarized below:

Table 3. Classification Report

	precision	recall	f1-score	support
Not Eligible	1.00	0.67	0.80	6
Eligible	0.92	1.00	0.96	24
accuracy			0.93	30
macro avg	0.96	0.83	0.88	30
weighted avg	0.94	0.93	0.93	30

1. Precision:. Precision measures the proportion of correct positive predictions out of all predicted positives.
   a. Not Eligible (0): Precision = 1.00, meaning all instances predicted as "Not Eligible" are correct.
   b. Eligible (1): Precision = 0.92, meaning 92% of the predictions classified as "Eligible" are indeed correct.
2. Recall: Recall measures the proportion of actual positives that are correctly identified.
   a. Not Eligible (0): Recall = 0.67, indicating that only 67% of the actual "Not Eligible" instances were correctly classified.
   b. Eligible (1): Recall = 1.00, meaning the model successfully identifies all "Eligible" instances as "Eligible."
3. F1-Score: The F1-Score is the harmonic mean of precision and recall, providing a balanced evaluation, especially in imbalanced datasets.
   a. Not Eligible (0): F1-Score = 0.80, reflecting moderate performance due to the lower recall.
   b. Eligible (1): F1-Score = 0.96, indicating excellent performance due to high precision and perfect recall.
4. Accuracy: Accuracy = 93%, meaning the model correctly predicts 93% of all instances. However, accuracy can sometimes be misleading, especially when class imbalance exists, as the model might perform well for the majority class (in this case, "Eligible") but poorly for the minority class

http://doi.org/10.26594/register.v8i2.XXX

ISSN 2460-7258 (online) | ISSN 1978-1520 (print)
7 (1) September 2024 285-294

Results

1. The model performs very well for the "Eligible" class (label 1) with high precision, recall, and F1-score. This shows that the model can effectively identify data that are truly eligible.
2. However, the model performs very poorly for the "Not Eligible" class (label 0), as it fails to predict any of the instances correctly (precision, recall, and F1-score for this class are 0).
3. Although the overall accuracy is 98%, this indicates that the model is heavily biased toward the "Eligible" class and fails to recognize the minority "Not Eligible" class. This is a common issue with imbalanced datasets.

3.3. Possible Improvements

1. Resampling Techniques: Implementing resampling strategies such as oversampling the minority class ("Not Eligible") or undersampling the majority class ("Eligible") to address the class imbalance in the dataset. These methods can help the model better learn patterns in the underrepresented class and improve recall for "Not Eligible" students.
2. Exploring Alternative Models: Using alternative models that are better suited for handling imbalanced datasets, such as Random Forest, Balanced Random Forest, or models that penalize misclassification of the minority class. These models can provide more robust performance by focusing on both “Eligible” and “Not Eligible” classifications.
3. Class Weight Adjustments: Applying class balancing techniques, such as setting class_weight='balanced' in the Decision Tree Model, to assign greater weight to the minority class ("Not Eligible"). This approach helps the model give more importance to correctly classifying "Not Eligible" students, reducing misclassification errors.

4. Conclusion

This research evaluated the Decision Tree Model for Classifying University Students Eligible for UKT Waivers. The model demonstrated strong performance overall, achieving a high accuracy of 93.33%, with excellent metrics for the "Eligible" class, including a precision of 92%, recall of 100%, and an F1-score of 96%. These results highlight the model's ability to effectively classify students eligible for UKT waivers.However, the performance for the "Not Eligible" class was less effective, with a lower recall of 67%, indicating that the model struggled to correctly identify all ineligible students. This disparity underscores the challenge of class imbalance within the dataset, as the majority of students belonged to the "Eligible" class. The confusion matrix further confirmed this issue, with two ineligible students misclassified as eligible.To address these limitations, future improvements could include the use of resampling techniques (e.g., oversampling the minority class or undersampling the majority class), employing alternative models like Random Forest or Balanced Random Forest, and applying class balancing strategies, such as assigning weights to the minority class. These approaches could enhance the model’s ability to classify "Not Eligible" students more accurately while maintaining strong performance for "Eligible" students.In conclusion, the Decision Tree Model provides a reliable and interpretable method for classifying UKT waiver eligibility, offering a valuable framework for decision-making in university financial aid allocation. Nevertheless, addressing class imbalance remains a key area for future to improve the model's overall fairness and robustness. Despite the limitations, the Decision Tree Model proved to be a highly interpretable and reliable tool for predicting UKT eligibility. Its ability to generate decision rules offers transparency, which is critical for educational institutions in justifying their financial aid allocation processes. The high performance for the "Eligible" class demonstrates the model's potential for implementation in real-world scenarios.

In conclusion, while the Decision Tree Model provides a robust framework for classifying UKT eligibility, addressing class imbalance is essential to further enhance its performance for the "Not Eligible" class. Future should focus on integrating the suggested techniques to improve the fairness and robustness of the model, ensuring that all students eligible and ineligible are classified accurately and equitably. This improvement will strengthen the model’s utility in supporting data-driven decisions in financial aid distribution.

 http://doi.org/10.26594/register.v8i2.XXX

ISSN 2460-7258 (online) | ISSN 1978-1520 (print)
7 (1) September 2024 285-294

5. Suggestion

Future on the Decision Tree model for classifying university students eligible for UKT waivers could focus on addressing class imbalance by applying techniques such as oversampling (e.g., SMOTE) or undersampling to improve recall for the ineligible class. Incorporating additional features, such as regional economic indicators, parental employment status, or attendance rates, may enhance model performance. The use of ensemble methods like Random Forest, Gradient Boosting, or XGBoost could also improve classification accuracy and robustness. Expanding the dataset to include more diverse student profiles across multiple universities would help generalize the model, while integrating explainability techniques such as SHAP (SHapley Additive exPlanations) could provide deeper insights into feature contributions, improving trust in decision-making. Comparing the Decision Tree with other algorithms, such as Logistic Regression or Neural Networks, might reveal trade-offs in performance and interpretability. Deploying the model in real-world settings and conducting cost-benefit analyses could justify its implementation by evaluating its financial and administrative efficiency over traditional processes. Additionally, incorporating temporal data to account for changes in eligibility status and exploring hybrid approaches that combine data-driven insights with domain expertise could further enhance accuracy and fairness in classifying students. These improvements would refine the application of machine learning in financial aid allocation, making the process more efficient, equitable, and scalable.

Furthermore, involving policymakers and university administrators in the model refinement process is essential to ensure that the system aligns with institutional goals and legal considerations. Feedback from stakeholders can guide the inclusion of relevant criteria and thresholds that reflect the values and priorities of the educational institution. This collaborative approach also supports the ethical implementation of artificial intelligence in decision-making.

Another important consideration is the continuous monitoring and validation of the model's performance after deployment. Regular audits, periodic retraining with updated datasets, and integration with real-time student information systems will help maintain the model's relevance and accuracy over time. This proactive maintenance ensures the system adapts to shifting socio-economic conditions and remains aligned with the evolving needs of students.

Finally, future implementations may benefit from incorporating user-centric design in system interfaces, allowing administrative staff to interact with the model easily and intuitively. Building a dashboard that visualizes decision paths, highlights key factors, and provides actionable recommendations would significantly enhance the practical utility of the model. By combining technical accuracy with usability and transparency, the system could play a pivotal role in modernizing financial aid processes across higher education institutions.

Moreover, ethical considerations must be integrated into the development and deployment of predictive models in educational contexts. Models that influence decisions on financial aid allocation should be evaluated not only in terms of performance metrics but also in terms of fairness, accountability, and transparency. Biases embedded in the training data whether due to socio-economic disparities, regional differences, or historical imbalances can inadvertently affect model outcomes. Therefore, applying fairness-aware machine learning techniques and conducting bias audits are vital steps to ensure equitable treatment of all applicants.

In addition, collaboration with interdisciplinary experts, including data scientists, education policy analysts, and social workers, can further enrich the model's development. Their diverse perspectives can help interpret the data more holistically and shape the design of the model to better reflect real-world complexities. Such interdisciplinary approaches also support the integration of contextual knowledge, which purely data-driven methods might miss.

To enhance public trust and acceptance, transparency about how the model works and how decisions are made is crucial. Creating accessible documentation, such as model cards or fact sheets, that explain the model's purpose, data sources, assumptions, and limitations can foster greater understanding among users, including students and guardians. Transparency initiatives like these not only improve stakeholder engagement but also serve as safeguards for ethical accountability.

Lastly, scalability and integration into existing institutional systems are practical aspects that must be planned carefully. As universities often use different platforms for student management and

 http://doi.org/10.26594/register.v8i2.XXX

ISSN 2460-7258 (online) | ISSN 1978-1520 (print)
7 (1) September 2024 285-294

financial services, ensuring interoperability with these systems is necessary for smooth implementation. Automating the eligibility assessment process using the model can significantly reduce administrative workload, minimize human error, and accelerate aid distribution timelines. However, careful planning is required to manage this transition, including training for staff, data privacy protection, and pilot testing before full-scale deployment.

References

[1] L. Rokach and O. Maimon, "Decision Trees," in Data Mining and Knowledge Discovery Handbook, 2nd ed., Springer, 2015, pp. 165–192.
[2] W.-Y. Loh, "Regression Trees and Forests for Classification," Annals of Data Science, vol. 6, no. 1, pp. 1–26, Mar. 2019, doi: 10.1007/s40745-018-0161-9.
[3] L. Breiman, Classification and Regression Trees, New York, NY, USA: Routledge, 2017.
[4] I. H. Witten, E. Frank, M. A. Hall, and C. J. Pal, Data Mining: Practical Machine Learning Tools and Techniques, 5th ed., Burlington, MA, USA: Morgan Kaufmann, 2018.
[5] J. Brownlee, Imbalanced Classification with Python: Better Metrics, Balance Skewed Classes, and Improve Model Performance, 1st ed., Melbourne, Australia: Machine Learning Mastery, 2018.
[6] R. E. Schapire, "Explaining AdaBoost," in Empirical Inference, Springer, 2015, pp. 37–52, doi: 10.1007/978-3-662-45620-3_2.
[7] A. Liaw and M. Wiener, "Classification and Regression by RandomForest," R News, vol. 2, no. 3, pp. 18–22, Dec. 2018.
[8] C. Zhang, Y. Wang, and J. Chen, "An Explainable Machine Learning Framework for Imbalanced Data Classification," IEEE Access, vol. 7, pp. 165841–165851, Nov. 2019, doi: 10.1109/ACCESS.2019.2952862.
[9] H. Liu, J. Wu, and F. Zhang, "Feature Selection and Classification Using Decision Tree Optimized with Ensemble Methods," Journal of Systems and Software, vol. 134, pp. 138–148, Apr. 2018, doi: 10.1016/j.jss.2017.10.027.
[10] F. Pedregosa et al., "Scikit-learn: Machine Learning in Python," Journal of Machine Learning , vol. 12, pp. 2825–2830, Oct. 2019.