BAREKENG: Journal of Mathematics and Its Applications
March 2025 Volume 19 Issue 1 Page 0665–0674
P-ISSN: 1978-7227 E-ISSN: 2615-3017
https://doi.org/10.30598/barekengvol19iss1pp0665-0674

IMPROVING CLUSTER ACCURACY IN TUITION FEES: A MULTILAYER PERCEPTRON NEURAL NETWORK AND RANDOM FOREST APPROACH

Sumin[1*], Prihantono[2], Khairawati[3]

[1]Mathematics Tadris Study Program, Faculty of Tarbiyah and Teacher Training, Institut Agama Islam Negeri Pontianak
[2]Master of Sharia Economics Study Program, Faculty of Tarbiyah and Teacher Training, Institut Agama Islam Negeri Pontianak
[3]Islamic Religious Education Study Program, Faculty of Tarbiyah and Teacher Training, Institut Agama Islam Negeri Pontianak
Jln. Letjen Soeprapto No. 19, Benua Melayu Darat, Pontianak, 78122, Indonesia.

Corresponding Author's e-mail: ***amien.ptk@gmail.com***

ABSTRACT

Article History:

Received: 25th August 2024
Revised: 29th November 2024
Accepted: 29th November 2024
Published: 13th January 2025

Keywords:

Artificial Neural Networks;
Single Tuition Fee;
Classification.
Random Forest

Manual classification of Single Tuition Fees (STF) has a high risk of misclassification due to the need for a more in-depth assessment of students' economic criteria. This research uses Artificial Neural Networks (ANN), specifically the Multilayer Perceptron (NN-MLP) model, to detect and correct errors in Single Tuition Fee (STF) classification. This study aims to apply the NN model to identify and correct classification errors in the STF clustering of State Islamic Religious Universities in Indonesia (PTKIN). This research was conducted using exploratory methods and quantitative approaches involving a population of PTKIN students throughout Indonesia. A sample of 282 respondents was selected using a simple random sampling method. The results showed that NN-MLP is an effective tool for identifying and correcting misclassification in determining PTKIN tuition fees with significantly improved classification accuracy characterized by an accuracy value of 71.28% and MSE of 0.287; this model can be used as a basis for developing information systems that are fairer and more accurate in managing tuition fees in higher education. This research also proves that the NN method is superior to traditional statistical methods and simple machine learning in handling complex and diverse data. In addition, the Random Forest model successfully identified the most influential input variables in STF classification. Father's occupation, mother's occupation, number of dependents, and utility bills such as water and electricity significantly contributed to the STF classification. In contrast, variables such as vehicle facilities showed a lower contribution.

How to cite this article:

Sumin, Prihantono and Khairawati., " IMPROVING CLUSTER ACCURACY IN TUITION FEES: A MULTILAYER PERCEPTRON NEURAL NETWORK AND RANDOM FOREST APPROACH," *BAREKENG: J. Math. & App.*, vol. 19, iss. 1, pp. 0665-0674, March, 2025.

Journal homepage: *https://ojs3.unpatti.ac.id/index.php/barekeng/*
Journal e-mail: *barekeng.math@yahoo.com; barekeng.journal@mail.unpatti.ac.id*
Research Article · Open Access

1. INTRODUCTION

Higher education in Indonesia has undergone many changes along with the times, one of which is the Single Tuition Fee (STF) policy. STF was introduced as an effort by the government to create fairness in the cost of education in public universities [1], in which the fees to be paid by students are adjusted to their or their parents' economic abilities [2], [3]. This system is expected to ensure that higher education is accessible to all levels of society, without burdening students from underprivileged families with unaffordable fees. However, despite these good intentions, the implementation of STF often leads to unforeseen problems. One of the problems that arise is the misclassification in the determination of students' STF group.[4], [5], [6].

This misclassification can be caused by various factors, such as inaccuracies in the assessment of student economic data or errors in filling out the form by the students themselves.[4], [6]. Careless or inaccurate form filling on the part of students can produce data that does not match reality, resulting in incorrect placement in certain STF groups. This inaccuracy can have a serious impact, especially for students who should be in the STF group with lower costs but are placed in the group with higher costs, or vice versa. This phenomenon indicates a misalignment between the initial objectives of the STF policy and its practical implementation in the field [7].

Since the issuance of the Minister of Research, Technology and Higher Education Regulation No. 39/2017, which regulates the distribution of STF based on economic ability, this system has experienced various challenges, especially related to the accuracy of the data used for classification. Various reports and studies have shown that many students feel unfair about their STF determination, which often does not reflect their actual economic means. Protests from students against the STF determination have occurred in various public universities, indicating that this misclassification problem is not trivial. This phenomenon demands more serious attention from academics and related parties to find a more accurate and fair solution to STF determination. [7], [8].

Previous research has tried to address this classification problem with various approaches, including traditional statistical methods and simple machine learning algorithms [9], [10], [11]. However, the results from these approaches are still inadequate, especially when faced with complex and diverse data such as student economic data. For example, logistic regression, which is often used for classification, often fails when there are non-linear relationships in the data. [12], while decision tree methods are prone to overfitting when faced with large and varied datasets [12], [13]. To address the existing practical gaps, we propose the use of a more sophisticated and adaptive method, namely Artificial Neural Networks (ANN), as a solution to improve classification accuracy in STF determination. NNs, particularly the Multilayer Perceptron (NN-MLP) model, have the ability to capture non-linear and complex patterns in data, thus allowing for more accurate and fair classification [14], [15], [16], [17], and Random Forest which is a novel method to rank variables based on their relevance to the classification problem, which can help reduce the number of model inputs in high-dimensional data sets and improve computational efficiency [18], [19]. The advantages of NN-MP and Random Forest in processing complex and diverse data make them promising alternatives to overcome the misclassification problem that often occurs in the STF system.

This research offers a combined application of JST-MLP and Random Forest to detect and correct misclassification in STF determination and categorize input variables based on their importance. This approach differs from previous research that has relied more on conventional statistical methods or simple machine learning algorithms. Combining the power of NN in complex data processing with the practical need for more accurate classification in STF settings, this research is expected to make a significant contribution in improving fairness and accuracy in tuition fee setting in public universities. The results of this study can also provide useful recommendations for relevant parties to address the problem of misclassification in the STF system more effectively. Based on the phenomena and research gaps that have been identified, the purpose of this study is to apply Artificial Neural Network models, specifically Multilayer Perceptron, to detect and correct misclassification in the Single Tuition Fee (STF) grouping.

2. RESEARCH METHODS

2.1 Research Designs, Approaches, and Types

This research uses exploratory research design, a quantitative approach, and survey research type. This method was chosen with the aim of demonstrating Artificial Neural Network modeling and Random Forest in detecting STF grouping errors and providing the best classification of the prediction results.

2.2 Research Variables

The variables used in this study consist of 10 independent variables and 1 dependent variable, which are Input Nodes (variables that affect the determination of a single tuition fee) are: Father's Income (X_1), Mother's Income (X_2), Father's Occupation (X_3), Mother's Occupation (X_4), Vehicle Facilities (X_5), Residential House Category (X_6), Home Ownership Status (X_7), Number of Dependents of Parents (X_8), Social Assistance from the Government (X_9), 3-month Electricity Bill (X_{10}), 3-month Water Bill (X_{11}), Classification of Single Tuition (Y).

2.3 Data Collection Methods and Instruments

The data collection method in this study was conducted using an online survey based on Google Forms, with a nominal scale questionnaire to facilitate data classification and analysis. The simple random sampling technique was chosen based on the relatively homogeneous characteristics of the population, namely students at State Islamic Universities in Indonesia, thus allowing simple random sampling techniques to achieve accurate and efficient representation. The sampling process begins with assigning a unique number to each registered member of the population; then, the sample selection is carried out randomly using a random number generator to avoid bias in the selection of respondents. Based on the statistical parameters, the confidence level was set at 90% ($Z = 1.64$), assuming a population proportion (p) of 0.5, representing the most conservative or maximum variation scenario. The margin of error was set at 4.88% ($d = 0.0488$), which was deemed adequate to produce generalizable conclusions with the desired confidence level. Given the very large or supposedly infinite population, the sample size (n) was calculated using the formula:

$$n = \frac{Z^2 . p(1-p)}{d^2} \tag{1}$$

Substitution of the available values resulted in the calculation: $n = \frac{1.64^2(0.5)0.5}{0.0488^2} \approx 282$. Based on these calculations, a sample size of approximately 282 respondents was obtained.

2.4 Data Analysis Tools

This research focuses on developing a Multi-Layer Perceptron Artificial Neural Network (NN-MP) and Random Forest approach to classify Single Tuition Fee (STF) categories based on variables such as parental income, employment, and home ownership status. This model is expected to provide a more precise and fair classification and ranking, in accordance with the financial condition of students. The research steps include data collection, processing, NN-MP modeling, training using the Backpropagation algorithm, evaluation, and interpretation of results.

2.5 Research Stages

First, data was collected through a survey of students that included the variables of parents' income, occupation, number of dependents, and home ownership status. After the data is collected, data normalization is performed to ensure that all variables are on the same scale, so that no variable dominates during the NN-MP training process. Data normalization is done with the formula:

$$x' = \frac{x - x_{min}}{x_{max} - x_{min}} \tag{2}$$

In **Equation (2)**, x' is the normalized data value, x is the original value of the data, x_{min} is the minimum value in the variable, and x_{max} is the maximum value. Normalization is very important to prevent the dominance of certain variables in the NN-MP training process and to accelerate the convergence of the model. **[20]**.

Secondly, NN-MP modeling is done by designing an architecture consisting of three main layers: input layer, hidden layer, and output layer. In the input layer, each variable in the survey data becomes one node. For example, if there are 11 variables, there will be 11 nodes in the input layer. The hidden layer is designed with 8 nodes, which are in charge of processing information from the input layer. The activation function used in the hidden layer is a sigmoid function, which converts linear inputs into non-linear outputs with the formula:

$$f(z) = \frac{1}{1+e^{-z}} \tag{3}$$

in **Equation (3)**, z represents the net input value of the nodes in the hidden layer. The sigmoid function allows the NN-MP to handle non-linear relationships between inputs and outputs, which is particularly useful in classification tasks [21].

Third, the training of the NN-MP model is initiated using the Backpropagation algorithm, which consists of two main stages: forward propagation and backward propagation. In the forward propagation stage, the input data is processed through the hidden layer until it reaches the output layer. The net input to the neurons in the hidden layer is calculated using the formula:

$$\mathrm{z}_{\mathrm{net_j}} = \sum_{i=1}^{n} (x_i . w_{ij}) + b_j \tag{4}$$

in **Equation (4)**, $\mathrm{z}_{\mathrm{net_j}}$ is the total input to the hidden neuron j, x_i is the input to the i, w_{ij} is the weight that connects the input x_i with the hidden neuron j and b_j is the bias added to adjust the calculation result. This formula explains that each input is multiplied by its weight, then summed, and the bias is added to produce the net input. The output of the hidden layer is then calculated using a sigmoid activation function and forwarded to the output layer to produce a STF category prediction. [22].

Fourth, in the backward propagation stage, the error is calculated as the difference between the predicted output and the actual target. The loss function used to calculate the error is the Mean Squared Error (MSE):

$$E = \frac{1}{2} \sum_{k=1}^{m} (d_k - y_k)^2 \tag{5}$$

In **Equation (5)**, E is the loss function, $d_k$is the target output for instance$k$instance, and y_k is the output predicted by the NN-MP. The MSE function measures the average of the squared differences between the prediction and the actual target, thus giving an idea of how good the model is at predicting the correct outcome. Once the error is calculated, the weights and biases in the model are updated using the gradient descent rule:

$$\Delta w_{ij} = -\eta \frac{\partial E}{\partial w_{ij}} \tag{6}$$

In **Equation (6)**, Δw_{ij}shows the change in the weights w_{ij}, η is the learning rate that controls how much changes at each iteration, and $\frac{\partial E}{\partial w_{ij}}$ is the derivative of the loss function against the weights. This update is done to minimize error and improve prediction accuracy in the next iteration [20].

Fifth, after the training is complete, the NN-MP model is evaluated to measure its classification accuracy. Accuracy is calculated using the formula:

$$\text{Accuracy} = \frac{\text{Number of Correct Predictions}}{\text{Total Precitions}} \tag{7}$$

In addition, the NN is evaluated using Mean Squared Error (MSE) to measure how well the model predicts the target value during the training process:

$$MSE = \frac{1}{m} \sum_{k=1}^{m} (d_k - y_k)^2 \tag{8}$$

Accuracy reflects how well the NN-MP model predicts the correct STF category, by comparing the number of correct predictions with the total predictions made. In addition to accuracy, the MSE value is also calculated to see how much error the model makes. A good model is characterized by high accuracy and low MSE, which indicates the model's ability to predict correctly and not deviate too much from the actual data. [23].

Sixth, the results of the NN-MP model are interpreted to understand the patterns found. Through the analysis of the results, it is possible to identify the main factors that influence the determination of STF categories, such as parental income or home ownership status. Based on these patterns, policy recommendations can be provided to adjust STF categories for students who may be misclassified. This interpretation is important to ensure that the results of the model are not only accurate but also meaningful in the context of broader applications. [21].

Seventh, the Random Forest method is used in the last stage of this research to rank the importance of input variables based on their contribution to prediction accuracy. First, the normalized data was divided into two parts: training data and testing data. This division ensures that the Random Forest model can be trained and tested effectively. Next, the Random Forest model is built using the training data by generating several decision trees. Each tree in Random Forest is built using a random subset of input variables and data examples, increasing the overall model's diversity and prediction accuracy [18]. Once the model is trained, the importance of variables is calculated using metrics such as Mean Decrease Impurity (MDI) or Mean Decrease Accuracy (MDA). For data that has correlations between variables, it is essential to use a variable importance measure that considers these correlations to avoid inaccurate results[24]. The formula for MDI is:

$$MDI = \sum_{t \in T} \frac{\Delta_i(t)}{|T|} \tag{9}$$

where $\Delta_i(t)$ is the decrease in impurity at node t, and $|T|$ is the number of trees in the Random Forest. This metric helps identify the variables that contribute the most to the model's predictions. The Random Forest model was then evaluated using test data with accuracy and Area Under Curve (AUC) metrics to measure performance in predicting the correct STF category. The use of AUC as the basis for calculating variable importance provides better results in situations of unbalanced data [25]. Finally, the variable importance ranking results are analyzed to identify the main factors that influence the assignment of STF categories. In the context of high-dimensional data, algorithms such as Boruta and Vita methods can improve the stability of variable selection [26]. This analysis provides a deep insight into the most influential variables, helping formulate more precise and data-driven policy recommendations.

The entire analysis process in this study was conducted using the open-source R software chosen for its ability to process complex data, including the Neural Networks feature that allows efficient NN-MP training and Random Forest. The use of Program R facilitates data normalization, model training with the Backpropagation algorithm, and evaluation of model results. The powerful statistical analysis features in Program R also facilitate the interpretation of results and the preparation of the final research report. Through Program R, the NN-MP training process becomes more structured, and the results obtained are more accurate and reliable for decision making.

3. RESULTS AND DISCUSSION

3.1 Misclassification Detection of STF Determination Results

We analyzed the data using the R program based on the steps outlined in the methodology section. First, we imported the research data containing input variables X_1 to X_{11} from an Excel file using the 'read_excel' function syntax. After the data was successfully imported, we calculated the total score for each student by summing the values of all input variables using the 'rowSums' function syntax. This total score is calculated as the cumulative contribution of each variable towards determining the ideal STF category. Next, we calculated the total score range by subtracting the minimum score from the maximum score. This range was then divided into five equal intervals. Where the range score is calculated as the difference between the maximum score and the minimum score. Based on these intervals, we grouped students into Ideal STF categories using mutate and case_when functions. Students with the highest total score were placed in the

STF 5 category, while those with the lowest score were placed in the STF 1 category. After calculating the STF Ideal category, we filtered the data to focus only on the STF Ideal and STF Realization variables, so other unnecessary variables were removed. The distribution of the Ideal STF category was examined using a function using the 'table' syntax. The results show that 52 students are in STF category 1, 127 are students in STF category 2, 74 students are in STF category 3, 22 students are in STF category 4, and 7 students are in STF category 5.

This distribution shows that the majority of students are in STF categories 2 and 3, which indicates a lower STF burden. In contrast, only a few students are in STF category 5, which reflects a higher STF burden. These results provide an initial picture of the ideal STF burden distribution based on the calculation of the total scores of the relevant input variables.

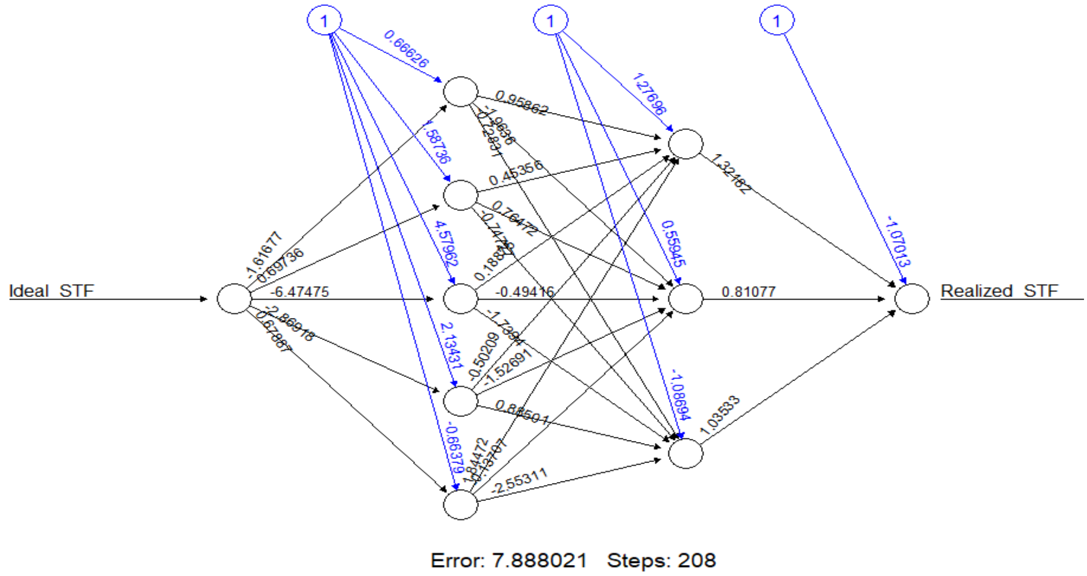

Figure 1. Artificial Neural Network Architecture

Figure 1 shows the structure of the neural network that has been trained to predict the Realized STF based on the Ideal STF. The network consists of one input node (Ideal STF), two hidden layers, and one output node (Realized STF). Each connecting line between the nodes in the network shows weights that reflect the strength of the relationship between the nodes. These weights, such as -6.47475 and 0.81077, are optimized during training to minimize the prediction error. The final error value achieved was 7.888021 after 208 training steps, which indicates the accuracy of the model in predicting the Realized STF from the Ideal STF. Overall, this artificial neural network successfully modeled the non-linear relationship between input and output through the hidden layer, and the relatively low error value indicates a fairly accurate prediction based on the trained data.

Table 1. Summary of STF Misclassification Detection Results

Sample	Ideal STF Group	Realized STF Group	STF Group Predicted by NN	Remarks
Student 01	2	2	2	True Classification
Student 02	1	4	2	Misclassification
Student 03	1	1	2	Misclassification
…	…	…	…	…
Student 282	4	3	3	True Classification

***Data Source**: Questionnaire processed using R software, Year 2024.*

The results of the analysis using the Multilayer Perceptron Artificial Neural Network (NN-MP) model show that the model successfully predicts the appropriate STF group for a number of students. Of the 282 students analyzed, there are 121 students whose classification matches between the STF realized by the university and the prediction generated by the model, which means that the model is able to identify the appropriate STF group for these students.

However, 161 cases were found where the model predictions differed from the realized STF. This discrepancy indicates that the STF is not by the student's condition as predicted by the model. The cases where predictions and realizations were not in the line indicate that the STF given to these students did not reflect the results that should have been, referring to the variables used in the model. This highlights the possibility of errors in STF assignments, where some students may be placed in inappropriate STF categories. This could have an impact on fairness in tuition allocation, as some students may pay more or less than they should based on a more objective analysis.

This analysis shows that the NN-MP model has a significant ability to detect misclassification in STF settings in higher education. The model is not only able to predict the correct STF group for some students but also identifies several cases where the STF setting may not be in accordance with the ideal conditions of students. The mismatch between the prediction and realization of STF in 161 cases indicates the need for evaluation and review in the STF setting process. The use of NN-MP can be an effective tool to support fairer and more informed decisions in tuition fee allocation, as well as ensuring that each student is placed in an STF group that is appropriate to their condition. This model offers a data-driven approach that can improve and refine STF setting policies in the future.

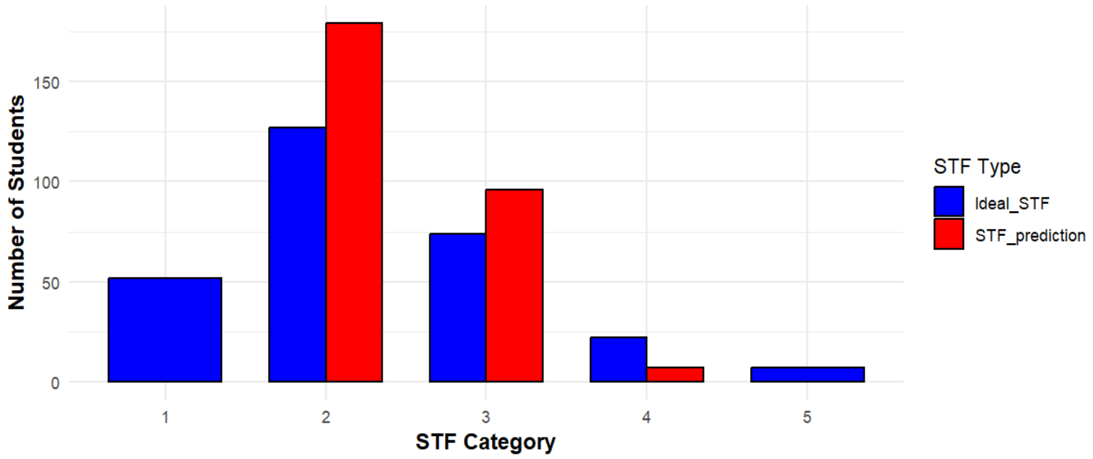

Figure 2. Distribution of Ideal STF and Predicted STF

The Ideal and Realized STF distribution graph in **Figure 2** shows the distribution of students based on the Ideal STF and Predicted STF categories generated by the Neural Network model. In STF Category 1, there are about 50 students, and most of them have predictions that match the Ideal STF, which is shown by the dominance of the blue color in this category. This indicates that the model is quite accurate in predicting the STF group for students in this category. The STF 2 category shows the largest number of students. The dominance of red at the bottom of the bar in Category 2 indicates that many students in this category have predictions that match the Ideal STF. However, the red top bar indicates that there are a number of students predicted by the model to be in Category 2, when their Ideal STF should be in Category 1 or Category 3, indicating misclassification. The STF 3 category also has a significant distribution, with the red color indicating a match between the Ideal STF and the prediction. However, the presence of red at the top of the bar indicates that there is a prediction error, where the model predicts some students to be in Category 3, even though their Ideal STF is in Category 2 or possibly even in Category 4. In STF Category 4, the number of students is smaller, and the proportion of prediction errors shown by the red color is smaller compared to Categories 2 and 3. This indicates that the Neural Network model has relatively better accuracy in this category. STF Category 5 has a very small number of students, and the graph shows that the model's predictions for this category are quite accurate, with little or no significant classification errors. In conclusion, although the Neural Network model can provide accurate predictions for many students, there are still some misclassifications, especially in Categories 2 and 3, indicating room for improvement in STF setting and prediction.

3.2 Prediction Accuracy

The Artificial Neural Network-Multilayer Perceptron (NN-MP) model in this study is applied to predict the Single Tuition Fee (STF) category based on various socio-economic input variables. The main purpose of this analysis is to evaluate the extent to which the NN-MP model can provide predictions in accordance with the Ideal STF category that has been calculated previously. Evaluation of prediction accuracy is done through the measurement of Mean Squared Error (MSE) and percentage accuracy as presented in Figure 3.

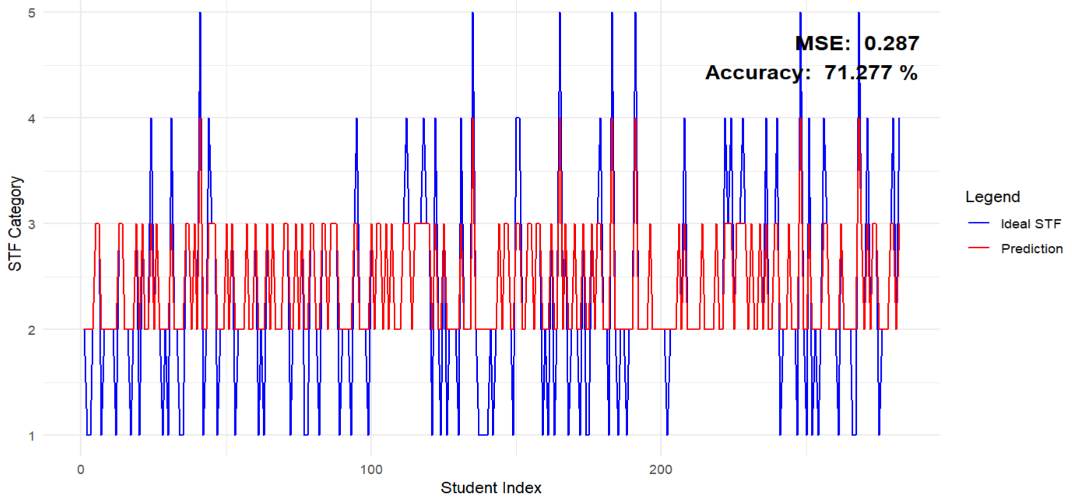

Figure 3. Prediction Accuracy

The graph in Figure 3 shows the comparison between the Ideal STF, represented by the blue line, and the Predicted STF, represented by the red line, for each student. The blue line is seen to have greater variation than the red line, indicating the tendency of the NN-MP model to produce more uniform predictions compared to the variation in the actual Ideal STF value. The model produced a Mean Squared Error (MSE) value of 0.29, indicating a moderate average squared difference between the Ideal STF and the Predicted STF. Although this MSE value is quite low, ideally a smaller value is required to indicate a minimal level of prediction error so that the prediction model can be considered more accurate. The NN-MP model achieved a prediction accuracy of 71.28%, indicating that the model managed to classify STF correctly in about 71% of cases. This accuracy indicates good model performance, although there is still room for improvement, especially regarding increasing the model's ability to minimize prediction errors. Model optimization or input variable adjustment may be required to further improve prediction accuracy.

Overall, the NN-MP model showed adequate capacity to predict the STF category, but the visible difference between the Ideal STF and the Predicted STF suggests the potential for further improvement. The accuracy of this prediction is crucial to ensure that the STF setting is in line with students' economic conditions, resulting in a fairer and more equitable system.

3.3 Input Variable Importance Analysis

In the next stage, further analysis was conducted to identify the variables that made the most significant contribution as the basis for STF classification. This process utilizes the Random Forest model, which is not only known to be effective in handling data complexity but can also be integrated or used as an extension of the Artificial Neural Network-Multilayer Perceptron (NN-MP) method. The combination of these two methods allows for a more comprehensive analysis, where NN-MP can be used to predict STF categories, while Random Forest helps understand which variables contribute most to the classification process.

Table 2. Level of Importance of Input Variables

Variables	Importance	Rating
Father's Occupation	52.25016	1
Mother's Occupation	26.69447	2
Number of Dependents	19.77231	3
Water Bill (3 months)	18.82896	4
Electricity Bill (3 months)	18.46363	5
Mother's Income	17.6954	6
Father's Income	12.23419	7
Government Social Assistance	11.87726	8
House Category	11.51351	9
House Ownership Status	11.09109	10
Vehicle Facility	1.168139	11

***Data Source**: Questionnaire processed using R software, Year 2024.*

The results of further analysis using the Random Forest model in **Table 2** provide insight into which variables contribute most to the STF classification based on their weight or importance. From the resulting table, the Father's Occupation variable has the highest importance weight of 52.25, making it the first-ranked variable. This shows that the father's occupation is the main factor that is most influential in determining the STF category. Mother's Occupation ranked second with an importance weight of 26.69, underscoring the importance of both parents' employment status in determining the family's economic ability to pay STF. Number of Dependents ranked third with an importance weight of 19.77, indicating that the number of family dependents is an important factor in STF classification. Meanwhile, Water Bill (3 months) and Electricity Bill (3 months) ranked fourth and fifth respectively with importance weights of 18.83 and 18.46, indicating that regular expenditure on household utilities is a significant factor. Parents' income is also important, with Mother's Income and Father's Income having importance weights of 17.70 and 12.23, ranking sixth and seventh, respectively. However, their contribution is lower compared to the employment-related variables and family burden. Other variables such as Government Social Assistance, House Category, and House Ownership Status make smaller contributions, with importance weights of 11.88, 11.51, and 11.09, respectively. Finally, Vehicle Facility has the lowest importance weight of 1.17, ranking last, which indicates that vehicle facilities are less relevant in decision-making regarding STF classification.

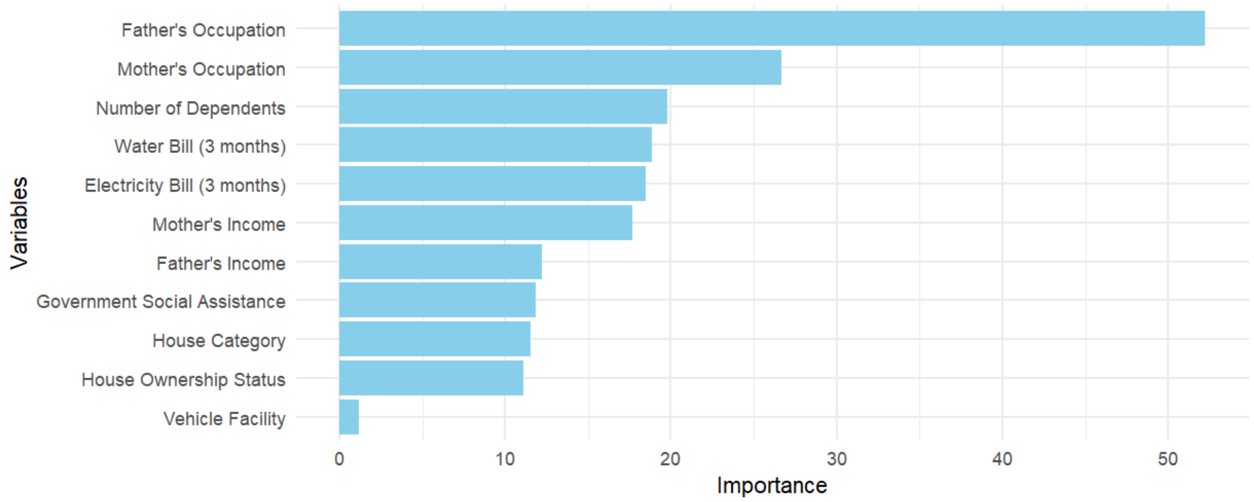

Figure 4. Random Forest Graph of Importance Level of Input Variables

The resulting graph in **Figure 4** confirms the results from **Table 2** by providing a clear visualization of the contribution of each variable in the Random Forest model. Father's Occupation stands out with the longest bar, indicating the highest contribution to the STF classification. Mother's Occupation, number of dependents, and utility bills also have significant bars, indicating the importance of these variables. In contrast, variables such as Vehicle Facility have the shortest bars, indicating the lowest contribution. This graph effectively illustrates the priority ranking of the variables, reinforcing the table's analysis of which variables most influence STF predictions.

4. CONCLUSIONS

The Neural Network-Multilayer Perceptron (NN-MP) and Random Forest methods effectively identified and corrected Single Tuition Fee (STF) misclassification at PTKIN. The significant increase in classification accuracy shows the great potential of NN-MP as the basis for developing a more fair and accurate information system in managing tuition fees. NN-MP excels at handling complex and diverse data and correcting classification errors with better results than traditional statistical methods and simple machine learning algorithms. Meanwhile, Random Forest successfully identified the most influential input variables in determining STF, such as the father's and mother's occupation, number of dependents, and utility bills (electricity and water). The vehicle facility variable was recorded to have the lowest contribution. This finding confirms the importance of family socio-economic variables in classifying tuition fees and offers a basis for more precise, fair, and data-driven fee management policies.

REFERENCES

[1] R. A. Yuliana and I. Hermawati, “Evaluasi kebijakan publik Universitas Negeri (PTN) terkait kebijakan kenaikan biaya kuliah,” *HUMANITIS: Jurnal Homaniora, Sosial dan Bisnis*, vol. 2, no. 7, pp. 614–621, 2024.
[2] A. N. Tsaqif, A. Setiawan, and F. N. Fajar, “Persepsi mahasiswa terhadap implementasi uang kuliah tunggal pada jurusan manajemen pendidikan Islam Fakultas Tarbiyah dan Keguruan Uin Alauddin Makassar,” *Educational Leadership: Jurnal Manajemen Pendidikan*, vol. 4, no. 1, pp. 42–58, 2024.
[3] R. Ayu, “Analisis Kebijakan Pada Peraturan Menteri Pendidikan No. 39 Tahun 2017 Tentang Biaya Kuliah Tunggal Dan Uang Kuliah Tunggal Pada Perguruan Tinggi Universitas Negeri Padang.,” *IJAM-EDU (Indonesian Journal of Administration and Management in Education)*, vol. 1, no. 3, pp. 194–209, 2024.
[4] W. N. Nursafitri and N. S. Ramadhan, “Analisis Klasifikasi STF Mahasiswa Berdasarkan Tingkat Penghasilan Orang Tua Menggunakan Algoritma C4. 5,” *Inventor: Jurnal Inovasi dan Tren Pendidikan Teknologi Informasi*, vol. 2, no. 1, pp. 1–9, 2024.
[5] Y. D. Retnoningsih and A. Marom, “Analisis Kebijakan Penyelenggaraan Pendidikan Berbasis Uang Kuliah Tunggal Bagi Perguruan Tinggi Negeri Fakultas Ilmu Sosial Dan Ilmu Politik Universitas Diponegoro Semarang Jawa Tengah,” *Journal of Public Policy and Management Review*, vol. 6, no. 2, pp. 482–497, 2017.
[6] S. Zaleha, “Problema penentuan uang kuliah tunggal: antara harapan dan kenyataan pada Institut Agama Islam Negeri Bengkulu,” -, 2022.
[7] M. Nasir, “Biaya Kuliah Tunggal Dan Uang Kuliah Tunggal Pada Perguruan Tinggi Negeri Di Lingkungan Kementerian Riset, Teknologi, Dan Pendidikan Tinggi,” Jakarta, Indonesia, 2017. [Online]. Available: kemenristekdikti.go.id
[8] M. Nasir, “Uang Kuliah Tunggal (STF) Wujud Keadilan Biaya Perkuliahan,” Jakarta, 2017. [Online]. Available: https://lldikti1.ristekdikti.go.id/details/apps/1806
[9] R. Efendi, D. Andreswari, and I. Barus, “Sistem Pendukung Keputusan Penentuan Uang Kuliah Tunggal Dengan Menggunakan Metode Weight Product,” in *SNIA (Seminar Nasional Informatika dan Aplikasinya)*, 2019, pp. 14–19.
[10] H. Kurniawan and S. Defit, “Data Mining Menggunakan Metode K-Means Clustering Untuk Menentukan Besaran Uang Kuliah Tunggal,” *Journal of Applied Computer Science and Technology*, vol. 1, no. 2, pp. 80–89, 2020.
[11] R. Andrea and N. Nursobah, “Penerapan Algoritma K-Medoids Untuk Pengelompokan Data Penerima Bantuan Uang Kuliah Tunggal Bagi Mahasiswa Terdampak Covid-19,” *Building of Informatics, Technology and Science (BITS)*, vol. 3, no. 4, pp. 632–638, 2022.
[12] E. Dumitrescu, S. Hué, C. Hurlin, and S. Tokpavi, “Machine learning for credit scoring: Improving logistic regression with non-linear decision-tree effects,” *Eur J Oper Res*, vol. 297, no. 3, pp. 1178–1192, 2022.
[13] Z. Nurizati, D. Vernanda, and T. Hendriawan, “Analisis Kelayakan Penurunan STF Pada Mahasiswa dengan Menggunakan Metode Decision Tree,” *Jurnal Tekno Kompak*, vol. 18, no. 1, pp. 90–100, 2024.
[14] M. Z. Abedin, C. Guotai, F. Moula, A. S. M. S. Azad, and M. S. U. Khan, “Topological applications of multilayer perceptrons and support vector machines in financial decision support systems,” *International Journal of Finance & Economics*, vol. 24, no. 1, pp. 474–507, 2019.
[15] A. Al Bataineh, D. Kaur, and S. M. J. Jalali, “Multi-layer perceptron training optimization using nature inspired computing,” *IEEE Access*, vol. 10, pp. 36963–36977, 2022.
[16] J. Naskath, G. Sivakamasundari, and A. A. S. Begum, “A study on different deep learning algorithms used in deep neural nets: MLP SOM and DBN,” *Wirel Pers Commun*, vol. 128, no. 4, pp. 2913–2936, 2023.
[17] R. A. Dunne, *A statistical approach to neural networks for pattern recognition*. John Wiley & Sons, 2007.
[18] A. Behnamian, K. Millard, S. N. Banks, L. White, M. Richardson, and J. Pasher, “A systematic approach for variable selection with random forests: achieving stable variable importance values,” *IEEE Geoscience and Remote Sensing Letters*, vol. 14, no. 11, pp. 1988–1992, 2017.
[19] A. Hapfelmeier, T. Hothorn, K. Ulm, and C. Strobl, “A new variable importance measure for random forests with missing data,” *Stat Comput*, vol. 24, pp. 21–34, 2014.
[20] I. Goodfellow, *Deep Learning*. MIT Press, 2016.
[21] S. Haykin, *Neural networks: a comprehensive foundation*. Prentice Hall PTR, 1998.
[22] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, “Learning representations by back-propagating errors,” *Nature*, vol. 323, no. 6088, pp. 533–536, 1986.
[23] Y. LeCun, Y. Bengio, and G. Hinton, “Deep learning,” *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.