REKAYASA
Journal of Science and Technology
https://journal.trunojoyo.ac.id/rekayasa

Rekayasa, 2025; 18(3): 418-429
ISSN: 0216-9495 (Print)
ISSN: 2502-5325 (Online)

Analysis of the Use of Random Forest Models to Measure the Quality of Indonesian Higher Education Institutions

Masdar Wiyono[1*], Cahyo Crysdian[1] , Mokhamad Amin Hariyadi[1] , Zainal Abidin[1],
Agung Teguh Wibowo Almais[1]
Faculty of Science and Technology Universitas Islam Negeri Maulana Malik Ibrahim
Jl Gajayana No 50 Kota Malang 65144 Jawa Timur
*E-mail Korespondensi : cahyo@ti.uin-malang.ac.id
DOI: https://doi.org/10.21107/rekayasa.v18i3.32024
Submitted July 24th 2024, Accepted November 30th 2025, Published December 7th 2025

Abstrak

Studi ini menyelidiki kinerja algoritma Random Forest dalam mengukur kualitas Institusi Pendidikan Tinggi (*Higher Education Institutions*/HEI) di Indonesia. Ketergantungan yang masih dominan pada proses penilaian administratif dan akreditasi konvensional sering kali dinilai kurang mampu menangkap performa institusi secara komprehensif, sehingga diperlukan pendekatan alternatif berbasis data. Penelitian ini mengusulkan pemanfaatan model Random Forest untuk melakukan klasifikasi tingkat kualitas institusi dengan mengacu pada indikator akreditasi yang relevan. Model RF-D menunjukkan performa klasifikasi yang optimal pada tiga kategori, yaitu Baik, Baik Sekali, dan Unggul, dengan nilai presisi, recall, dan F1-score yang tinggi pada setiap kelas. Kategori Baik Sekali mencapai presisi sebesar 89% dan recall 80%, sementara kategori Unggul memperoleh recall tertinggi sebesar 86%. Selain itu, skor Area Under Curve (AUC) yang mendekati 1,0 pada seluruh kategori memperkuat temuan mengenai kemampuan diskriminatif model yang sangat baik. Penelitian ini juga menyoroti pengaruh rasio pembagian data train-test terhadap stabilitas model. Ketidakseimbangan rasio yang terlalu ekstrem terbukti dapat memicu overfitting atau underfitting, sehingga pemilihan konfigurasi yang tepat sangat penting dalam pengembangan model. Secara keseluruhan, hasil penelitian ini menunjukkan bahwa Random Forest, ketika dikonfigurasi dengan parameter yang optimal, menawarkan pendekatan yang lebih akurat, objektif, dan dapat direplikasi dalam mengevaluasi kualitas HEI di Indonesia. Temuan ini diharapkan dapat berkontribusi pada pengembangan sistem asesmen pendidikan tinggi yang lebih transparan serta mendukung pengambilan keputusan berbasis data oleh pemangku kebijakan.

Kata Kunci: *random forest,* kualitas perguruan tinggi, evaluasi kinerja, *machine learning,* klasifikasi

Abstract

This study investigates the performance of the Random Forest algorithm in measuring the quality of Higher Education Institutions (HEIs) in Indonesia. The current reliance on administrative evaluations and conventional accreditation processes often fails to capture the institutions' actual performance comprehensively, indicating the need for a data-driven alternative. This research proposes the use of a Random Forest–based classification model to assess institutional quality based on relevant accreditation indicators. The RF-D model demonstrates optimal classification performance across three quality categories—Good, Very Good, and Excellent—with high precision, recall, and F1-scores for all classes. The Very Good category achieves a precision of 89% and a recall of 80%, while the Excellent category records the highest recall at 86%. Furthermore, the Area Under Curve (AUC) scores, which approach 1.0 for all categories, confirm the strong discriminative capability of the model. This study also highlights the influence of train–test data ratios on model stability. Extreme imbalances in data splitting can lead to overfitting or underfitting, emphasizing the importance of selecting an appropriate configuration during model development. Overall, the findings indicate that Random Forest, when optimized with suitable parameters, provides a more accurate, objective, and replicable approach for evaluating HEI quality in Indonesia. These results are expected to contribute to the development of a more transparent higher education assessment system and support data-driven decision-making among policymakers.

Key words*: random forest, higher education quality, performance evaluation, machine learning. classification*

INTRODUCTION

Higher education is a post-secondary education level that aims to equip students with academic and professional competencies to become members of society. College graduates are expected to be able to apply, develop, and create knowledge in science, technology, and the arts (Karim, 2020). Currently, the development of higher education in Indonesia continues to increase (Nulhaqim et al., 2016). This is evident in the increasing number of new HEIs each year, including universities, academies, institutes, polytechnics,

and colleges, both public and private. According to PDDIKTI data from 2023, there are a total of 4,437 HEIs in Indonesia (Kemdikbudristek, 2023).

According to Law Number 12 of 2012, a higher education institution is defined as an educational institution that provides higher education. In Indonesia, higher education is divided into two types: state universities, established and managed by the government, and private universities, run by the community through foundations or other organizing bodies. Referring to Law Number 20 of 2003 and Government Regulation Number 60 of 1999 concerning Higher Education, universities, both through their faculties and study programs, have the main task of developing and improving the quality of human resources to support the progress of the nation and civilization.

The increasing number of HEIs in Indonesia, has led to increasingly fierce competition in terms of educational quality. To ensure optimal educational quality, the government has established national higher education standards covering aspects of education, research, and community service (Pramana et al., 2021). However, challenges arise in implementing and measuring these standards, particularly in ensuring that each university meets the established criteria. The lack of an effective evaluation system and strict oversight mechanisms can lead to disparities in quality across universities (Sampe, N. & Arifin, Z., 2024). Furthermore, measuring the quality of HEIs in Indonesia still relies heavily on administrative evaluation methods and conventional accreditation processes, which, while useful, often do not fully reflect an institution's actual performance. Such evaluations tend to be periodic, static, and lack the flexibility to capture the dynamics of educational quality in real time.

Therefore, an objective, transparent, and sustainable measurement method is needed to assess the quality of HEIs in Indonesia to ensure the quality of higher education and support the achievement of national education goals. Most previous research in Indonesia has used only one algorithm or focused on a single aspect of quality, such as accreditation or graduate perceptions, without comparing the performance of two different algorithms. Few studies have directly examined the use of machine learning to analyze the quality of state universities holistically and based on actual data (Utami, SGA., 2025).

The method for measuring the quality of HEIs in Indonesia can be carried out through a data mining approach, which allows for in-depth analysis of various aspects such as academic performance, student satisfaction, graduate retention, and research productivity (Rohmawati et al., 2015; Romero & Ventura, 2007). Techniques such as classification and association can reveal hidden patterns in the data. Various models can be used, such as Linear Regression, Decision Tree, SVM, Logistic Regression, and Random Forest. Based on a study by Estrada Real & Cantu-Ortiz, the Random Forest algorithm proved to be one of the most accurate, with an accuracy rate of 99.48% in classifying universities into the Top 100 and Top 200 groups in the QS-WUR. Therefore, Random Forest is the primary choice for measuring higher education quality due to its ability to handle complex data and provide highly accurate prediction results (Estrada-Real & Cantu-Ortiz, 2022).

The concept of the Random Forest model is based on the Ensemble Learning principle, which combines multiple decision trees to produce more accurate and stable classifications. This model has proven effective in handling large, complex data sets with various feature types. The advantages of Random Forest lie in its ability to handle outliers and imbalanced data, as well as reduce the risk of overfitting. However, its disadvantages include a relatively longer training time and a lack of simple interpretation compared to linear models (Breiman, 2001).

The research further explores dataset limitations and the need for model enhancement, including feature and parameter optimization, to strengthen accuracy and generalization in the context of quality assessment for higher education institutions. Similar methodological approaches have been recommended in international studies on evaluating higher education quality using machine learning and learning analytics (Fidalgo-Blanco et al, 2023; Siemens & Long, 2011). This effort aligns with prior research highlighting the importance of comprehensive and continuous modeling within higher education systems (Supriyadi et al, 2022; Rohman et al., 2025). Accordingly, this data-driven approach is expected to provide a more holistic and equitable depiction of institutional performance and contribution to education and society.

The findings of this study are expected to offer strategic recommendations for stakeholders in improving the quality of higher education in Indonesia. Moreover, a deeper understanding of the key

aspects of quality assessment may support the development of more effective and sustainable higher education policies, in line with international practices in data-informed education (Papamitsiou & Economides, 2014)

RESEARCH METHOD

This study employs a quantitative approach, as the researcher utilizes numerical data analyzed through statistical methods, specifically the Confusion Matrix and the Area Under the Curve (AUC) score (Queirós et al., 2017; Ponto, 2015). In general, quantitative content analysis is a scientific research technique aimed at identifying characteristics and drawing conclusions from a body of content. This method is designed to systematically identify communication patterns in an objective, valid, reliable, and applicable manner (Han et al., 2011). In Figure 1, this study was conducted sequentially following eight main stages: literature review, data collection, data preparation, system design, system implementation, experimentation, model analysis, and conclusion drawing. These stages align with the methodological recommendations for machine learning research (Kotsiantis, 2007; Witten et al., 2011).

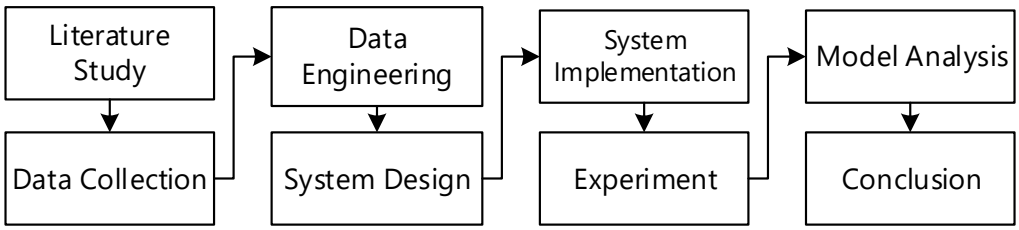


Figure 1. Research Procedure

Data Collection

Second, the data collection process involved collecting a dataset of higher education quality indicators from the PDDIKTI (https://pddikti.kemdiktisaintek.go.id/) or Higher Education Database and the SINTA website (https://sinta.kemdikbud.go.id/affiliations) to meet one of the higher education quality indicators. This dataset consists of 10 attributes, presented in Table 1.

Table 1. Indonesian Higher Education Quality Indicator Dataset

Attribute	Data Range	Data Type	Description
University	-	Categorical	Name of the higher education institution.
KPI_Score	9.2975 to 113.3688	Numerical	Score from the main performance indicators processed by the researcher.
Study_Programs	4 to 310	Numerical	Number of study programs at the university.
New_Students	39 to 183,905	Numerical	Number of newly enrolled students at the university.
Enrolled_Students	36 to 1,205,711	Numerical	Number of students currently enrolled in courses.
Graduates	0 to 77,881	Numerical	Number of graduates from the university.
Lecturers	0 to 3,066	Numerical	Number of lecturers at the university.
Sinta_Score_3Yr	593 to 1,230,039	Numerical	Publication score over the last 3 years for the university.
Sinta_Score_Overall	721 to 4,280,397	Numerical	Total cumulative publication score for the university.
Accreditation	Good to Excellent	Categorical	Accreditation category of the university, used as the target variable.

The Table 1 above explains the various attributes or variables used in this study to measure the quality of HEIs in Indonesia using the Random Forest model. The "University" attribute represents the name of each state university studied. This attribute is categorical because it is used solely as an institutional identifier, not as a numerical analysis variable. Furthermore, the KPI_Score, or Key Performance Indicator score, is an important variable that describes the level of success of HEIs in achieving the targets set by the Ministry of Education (Fitrah, 2018). This score was obtained from the researcher's processing of key indicator data, with values ranging from approximately 9.29 to 113.36. The higher this score, the better the institution's achievements in terms of governance and quality of education (Asnawi, 2005).

Each university offers programs ranging from 4 to 310. This number reflects the scope of an institution's academic fields the more study programs, the broader the academic spectrum offered to students (Estrada-Real & Cantu-Ortiz, 2022). Meanwhile, the New_Students and Enrolled_Students attributes indicate the number of new students accepted and the total number of students currently studying at the university, respectively. With a fairly large numerical range (from tens to hundreds of thousands), these variables provide information about the institution's attractiveness and capacity to accept and retain students (Romero & Ventura, 2020; Shahiri et al., 2015).

Furthermore, the Graduates attribute reflects the number of graduates produced by an institution in a given time period. This number is also an important indicator in assessing the institution's contribution to providing an educated workforce in society (Angkotasan & Watianan, 2021). Meanwhile, Lecturers indicates the number of lecturers a university has. This variable is crucial because the presence and qualifications of lecturers significantly influence the quality of learning, research, and community service. Furthermore, the Sinta_Score_3Yr and Sinta_Score_Overall attributes represent the university's research achievements and scientific publications. This score is calculated based on the number of lecturers' publications in the last three years and overall, according to data from the SINTA (Science and Technology Index) platform. A high score indicates a university's productivity and contribution to the development of science (Syahrir et al., 2023). Finally, Accreditation is the target variable in this study. This attribute is categorical and indicates the university's accreditation status, ranging from "Good" to "Excellent." This accreditation represents formal recognition of the institution's quality by the Indonesian higher education quality assurance agency (Mitra, 2023).

Data Engineering

Third, in the data preparation stage, the author prepares the data by cleaning it according to the required dataset specifications. This step aims to prevent overfitting and to increase model accuracy when measuring quality using the Min–Max Scaling technique (Sinsomboonthong, 2022). Min–Max Scaling has several characteristics: First, it maintains the relative relationship between original values, allowing the inherent patterns in the data to remain intact (Patro & Sahu, 2015). Second, if the normalized data fall outside the expected range, Min–Max Scaling can help detect inconsistencies or anomalies in the dataset (Han, Kamber, & Pei, 2012). Third, the main limitation of Min–Max Scaling is that it is highly sensitive to outliers, which can distort the scaled range and produce unrepresentative values (Jain et al., 2005).

$$x_{norm} \frac{x - \min(x)}{\max(x) - \min(x)} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (1)$$

Where x is normalized data, $xnorm$ is normalized data, min(x) is smallest value of all data, and max(x) is largest value of all data.

System Design

Fourth, the system was designed by constructing a data processing flow, from input to modeling using the Random Forest algorithm. This design included selecting important features and adjusting model parameters to suit the data characteristics (Figure 2). In this system, the process begins with collecting data on quality indicators for Indonesian HEIs, such as data attributes. This data serves as the primary basis for the analysis. Data preprocessing is then performed, namely the stage of cleaning and transforming the data to prepare it for analysis by the model. This includes handling missing values (Little & Rubin, 2019), normalization, and data type conversion.

The processed data is then analyzed using the Random Forest algorithm, which works by building multiple decision trees to generate accurate and stable predictions of higher education quality. Random Forest was chosen for its ability to process complex data and avoid overfitting (Cutler et al., 2012). Finally, the system produces output in the form of HEIs' quality predictions in the categories of Good, Very Good, or Excellent. These results are expected to assist policymakers in improving the quality of higher education strategically and sustainably.

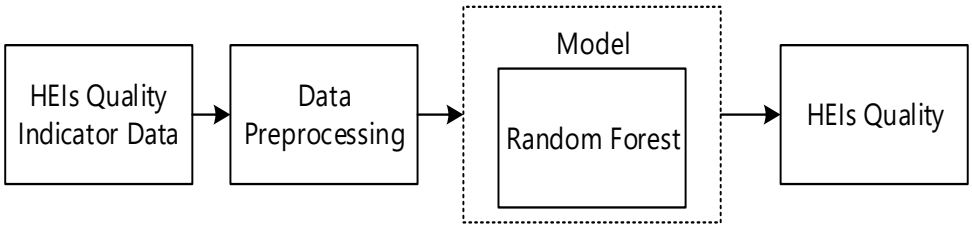


Figure 2. Design System

System Implementation

Fifth, the system is implemented through the Random Forest algorithm using the Python-based Kaggle Editor. The model is trained using training data and tested using pre-shared test data, allowing for objective measurement of model performance. Each Random Forest model consists of a set of decision trees randomly constructed using a subset of the training data through bootstrap sampling. Furthermore, each tree utilizes only a random subset of the available features at each branch (node), aiming to increase diversity between trees and reduce the risk of overfitting (Ying, 2019). In the context of university quality classification, each tree provides a categorical prediction for the class as the target variable. The system then calculates the final result based on the majority vote from all trees in the ensemble (2). This model is effective because it combines predictions from multiple weak models into a single, stronger model, which is more stable and accurate.

$$\hat{y} = mode\ \{h_1(x), h_2(x), .., h_n(x)\} \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots (2)$$

Where: h_i is the prediction of the tree th-i to input x, n is the number of trees in model, and mode is the largest value of all tree predictions.

Experiment

Sixth, experiments were conducted to test the model under various parameter configurations and scenarios. The goal was to evaluate the model's ability to accurately and stably predict university quality. Table 2 shows the parameter configurations used in this study, where the number of decision trees (n_estimators) was set at 300. This number was considered sufficient to form a stable and accurate model ensemble (Dudek, 2022). Furthermore, the maximum tree depth (max_depth) was limited to 10 levels to avoid overly deep and complex trees. The min_samples_split parameter was set to 5, meaning each node must have at least five samples to be split, and the min_samples_leaf value was set to 1, meaning each leaf in the decision tree contains at least one data point. The class_weight parameter was assigned a value of 'balanced', so class weights were automatically adjusted based on the proportion of occurrences of each class. Finally, the random_state parameter was set to 42 to ensure consistency of the model's results (Lotfi et al., 2025).

Table 2. Random Forest Model Parameter Configuration

Parameter	Value	Description
n_estimators	300	Number of trees in the forest
max_depth	10	Maximum depth of each decision tree
min_samples_split	5	Minimum number of samples required to split an internal node
min_samples_leaf	1	Minimum number of samples required at a leaf node
class_weight	'balanced'	Class weights are adjusted to handle imbalanced data
random_state	42	Random seed to ensure reproducibility

Table 3 shows the experimental research scenario with a composition of 118 data in total for each attribute divided into five groups with the composition of training data and test data of the RF-A model of

approximately 90%: 10%, RF-B models of approximately 80%: 20%, RF-C models of approximately 70%: 30%, RF-D models of approximately 60%: 40%, and RF-E models of approximately 50%: 50%.

Table 3. Experimental Scenario

Model	Code	Data Ratio	Training Data	Test Data	Model Name
Random Forest	A	90:10	106	12	RF-A
	B	80:20	94	24	RF-B
	C	70:30	82	36	RF-C
	D	60:40	70	48	RF-D
	E	50:50	58	60	RF-E

Model Analysis

Seventh, model analysis was conducted using evaluation metrics such as accuracy, precision, recall, F1-score, AUC score calculated using the Confusion Matrix, and interpretation of the AUC value (Table 4). By combining the results of the Confusion Matrix and the AUC score, the assessment of model performance in this study becomes more comprehensive and accurate, and can be used as a basis for selecting the best model for the predictions made (Muschelli III, 2020). The results of this analysis serve as the basis for determining the effectiveness of the Random Forest model in classifying the quality of HEIs in Indonesia.

Tabel 4. Confusion Matrix and Interpretation of AUC Values

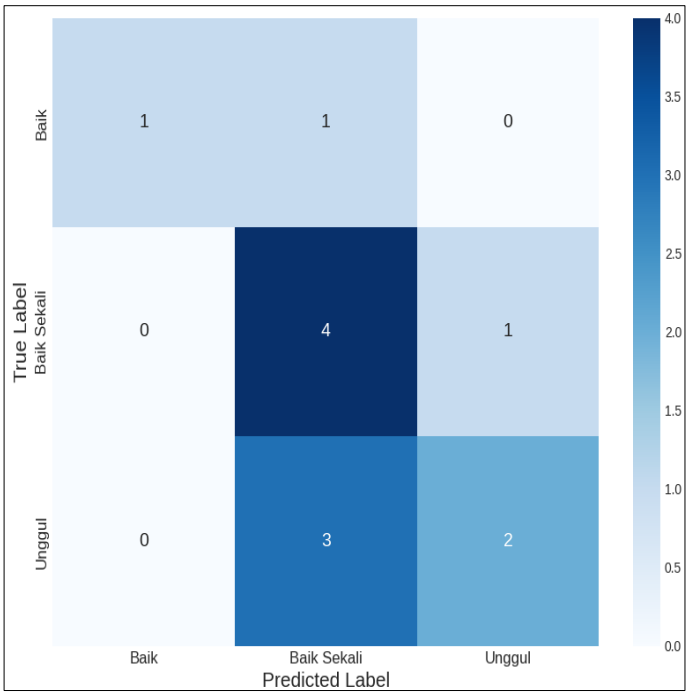
Predicted Class	Actual Positive Class	Actual Negative Class
Positive	TP (True Positive)	FP (False Positive)
Negative	FN (False Negative)	TN (True Negative)
AUC Score		
AUC Value	**Interpretation**	
0.90 – 1.00	Excellent	
0.80 – 0.90	Good	
0.70 – 0.80	Fair	
0.60 – 0.70	Poor	
0.50 – 0.60	Very Poor / No Better Than Random	

RESULTS AND DISCUSSIONS

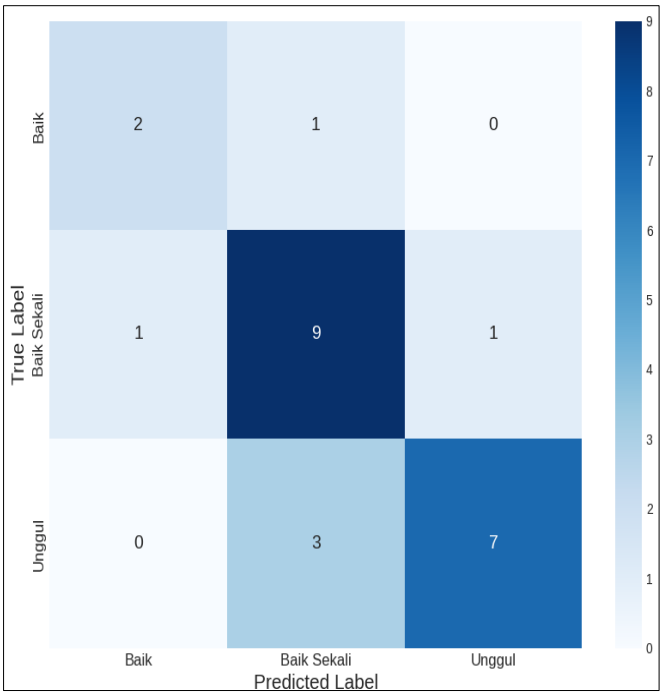
The Random Forest model in this study was designed to solve a classification problem by determining three quality categories for HEIs in Indonesia: "Excellent," "Very Good," and "Good" based on various indicators. This model was chosen due to its superior ability to manage large-scale data with many variables and its effectiveness in handling complex classifications (Ramadhan et al., 2021).

The process of measuring the quality of HEIs in Indonesia using the Random Forest model involves dividing the initial data into two main parts: training data and testing data based on experimental scenarios (Noviyanto et al., 2024). This division aims to ensure that the model can learn optimally from a subset of the data, and then test its accuracy using other, previously unseen data. This is crucial so that the model's performance reflects its ability to handle new, real-world data.

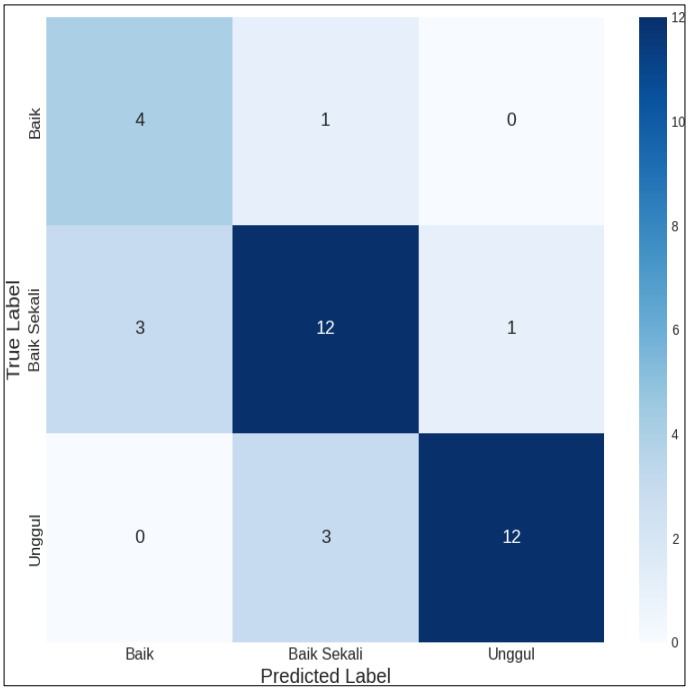
Next, to explore how data partitioning affects training results, several variations in the proportion of training and testing data were used. Each variation is assigned a specific code: RF-A, RF-B, RF-C, RF-D, and RF-E, each indicating a difference in the proportion of data partitioning. These models were then trained to measure the quality of higher education institutions based on the level of Accreditation, which is the target variable of the research, with three Accreditation categories, namely Good, Very Good, and Excellent.


True Label
Baik
Baik Sekali
Unggul
1 1 0
0 4 1
0 3 2
Baik
Baik Sekali
Unggul
Predicted Label
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0

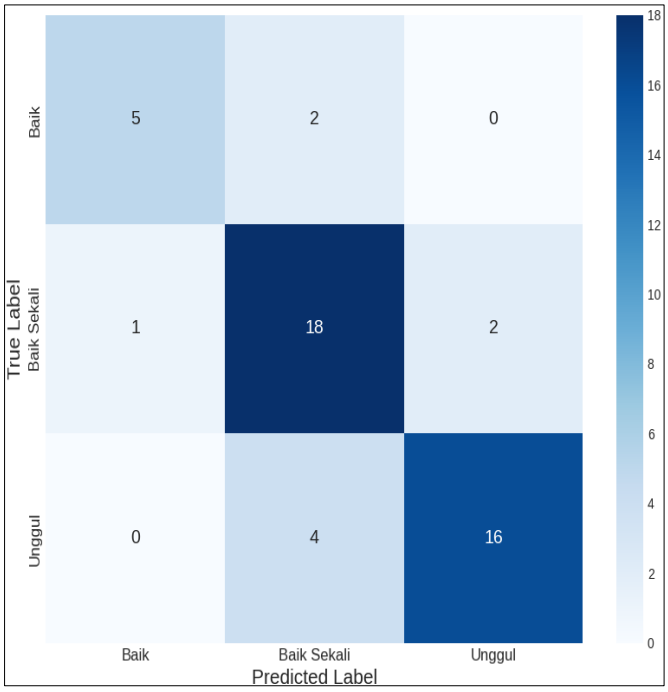
(a)


True Label
Baik
Baik Sekali
Unggul
2 1 0
1 9 1
0 3 7
Baik
Baik Sekali
Unggul
Predicted Label
9
8
7
6
5
4
3
2
1
0

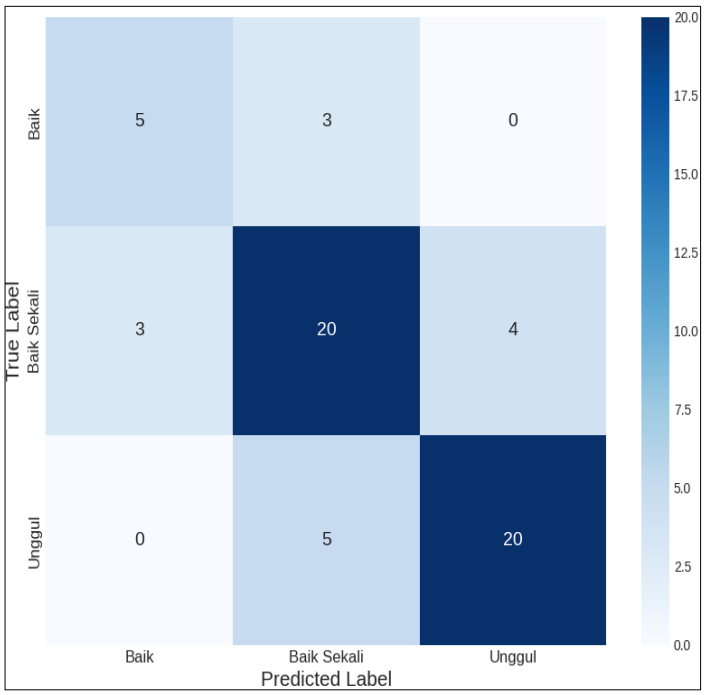
(b)


True Label
Baik
Baik Sekali
Unggul
4 1 0
3 12 1
0 3 12
Baik
Baik Sekali
Unggul
Predicted Label
12
10
8
6
4
2
0

(c)


True Label
Baik
Baik Sekali
Unggul
5 2 0
1 18 2
0 4 16
Baik
Baik Sekali
Unggul
Predicted Label
18
16
14
12
10
8
6
4
2
0

(d)



(e)

Figure 3. Confusion Matrix Random Forest Model Measurement Results: (a) RF-A, (b) RF-B, (c) RF-C, (d) RF-D, and (e) RF-E

Based on the experiments conducted in this study, Figure 3 (a-e) shows that the RF-A model's classification ability for each HEIs quality category is uneven. The model performed quite well in recognizing the Very Good category, with a success rate of four out of five data points, but was still less than optimal in distinguishing between the Good and Excellent categories. In the Good category, only one out of two data points were correctly classified, while in the Excellent category, only two out of five data points were correctly predicted. This indicates that the model still struggles to recognize the characteristics of the lowest and highest categories.

Furthermore, the RF-B model's measurement results showed fairly stable performance, particularly in recognizing the Very Good category, with a success rate of nine out of eleven data points. However, its classification ability for the Good and Excellent categories was still not fully optimal. For the Good category, two out of three data points were correctly classified, while one was incorrectly predicted as Very Good. Meanwhile, in the Excellent category, only seven out of ten data points were correctly predicted, with the remainder classified as Very Good. This indicates that the model still faces challenges in clearly distinguishing the characteristics of the highest and lowest quality categories.

Meanwhile, the RF-C model's measurement results showed quite good overall performance. The model successfully recognized the Good category with high accuracy, with four out of five data points correctly predicted. In the Very Good category, the model performed dominantly, with a success rate of twelve out of sixteen data points, although there were still minor misclassifications into the Good and Excellent categories. Meanwhile, for the Excellent category, the model demonstrated strong classification ability, with twelve out of fifteen data points correctly classified. These results indicate that the RF-C model is quite reliable in distinguishing the quality of higher education institutions, especially in the Very Good and Excellent categories, although there is still slight overlap between category features.

Furthermore, the RF-D model demonstrated fairly stable measurement performance. In the Good category, the model successfully classified five out of seven data points correctly, although two were still misclassified as Very Good. For the Very Good category, the model performed dominantly, with eighteen out of twenty-one data points correctly classified, while a small number were misclassified into other classes. In the Excellent category, sixteen out of twenty data points were accurately predicted, with four

others being misclassified as Very Good. These results indicate that the RF-D model has high sensitivity, especially in the intermediate category (Very Good), although there are still challenges in distinguishing the characteristic boundaries between the Excellent and Excellent categories.

Similarly, the RF-E model demonstrated quite good measurement performance. In the Excellent category, five of eight data sets were correctly classified, although three were misidentified as Excellent. In the Excellent category, the model performed predominantly, with twenty of twenty-seven data sets accurately classified, while the remaining were incorrectly classified as either the upper or lower class. In the Excellent category, twenty of twenty-five data sets were correctly predicted, while the remaining five were incorrectly classified as Excellent. These results indicate that the RF-E model is quite reliable in recognizing patterns in the Excellent and Excellent categories, although there are still challenges in distinguishing the boundaries between classes, particularly between the intermediate and highest categories.

Table 5. Random Forest Model Performance

Model Name	Data Ratio	Accuracy Value
RF-A	90:10	58.33%
RF-B	80:20	75%
RF-C	70:30	77.78%
RF-D	**60:40**	**81.25%**
RF-E	50:50	75%

Table 5 shows the model performance in measuring the quality of HEIs in Indonesia. The RF-A model, with a 90:10 training-to-test ratio, demonstrated the lowest accuracy of 58.33%, likely due to overfitting due to insufficient test data. The RF-B (80:20) model recorded a significant improvement with an accuracy of 75.00%, while the RF-C (70:30) model achieved the highest accuracy of 77.78%, indicating an optimal balance between training and testing. Although the RF-D (60:40) achieved an accuracy of 81.25%, its classification stability requires further evaluation. The RF-E (50:50) model achieved an accuracy of 75.00%, providing a fair and realistic classification distribution, although at risk of underfitting.

Overall, the RF-D model performed best in terms of accuracy with an accuracy of approximately 81.25%, followed by the RF-C model with an accuracy of approximately 77.78%, and the RF-B and RF-E models each with an accuracy of approximately 75%. Meanwhile, RF-A was the lowest-performing model with an accuracy of around 58.33%. This finding demonstrates the importance of selecting the appropriate training-to-test data ratio in building a reliable classification model for measuring the quality of HEIs. Furthermore, it is crucial to consider per-category evaluation and the likelihood of classification bias, such as middle bias, to produce a more accurate and fair prediction system. Middle bias typically occurs when the distinction between successive classes is not sharp enough or the predictive features are not strong enough to differentiate between classes (Hernández-Lobato et al., 2014).

Furthermore, to understand the model's performance in measuring the quality of HEIs in Indonesia, this study conducted an in-depth analysis of the best model, the RF-D, based on evaluation indicators such as precision, recall, F1-score, TP, TN, FP, FN, and AUC scores for each target variable class.

Table 6. In-depth Analysis of the Best Random Forest Model (RF-D)

Category	Precision (%)	Recall (%)	F1-Score (%)	TP	TN	FP	FN
Good	83	71	77	5	40	1	2
Very Good	75	86	80	18	21	6	3
Excellent	89	80	84	16	26	2	4

The in-depth analysis of the best model, Random Forest RF-D, in Table 6, shows that it performs quite well in measuring the quality of HEIs. In the Good category, the model recorded a precision of 83%, a recall of 71%, and an F1-score of 77%, with 5 correctly classified data (TP), but there were still 2 FN and 1 FP, indicating errors despite relatively high accuracy. For the Very Good category, the model had a precision

of 75%, a recall of 86%, and an F1-score of 80%, with 18 TP, 6 FP, and 3 FN, indicating stable performance despite a tendency to predict this class when the data is ambiguous. Meanwhile, the Excellent category showed the most solid performance, with a precision of 89%, a recall of 80%, and an F1-score of 84%, as well as 16 TP, 2 FP, and 4 FN, illustrating the model's high ability to recognize the highest category.

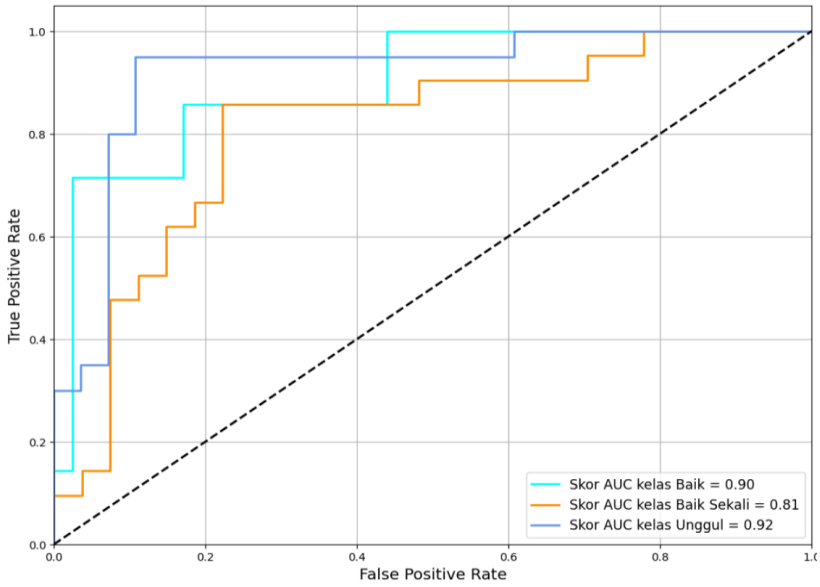


Figure 4. AUC Score of the Best Random Forest Model (RF-D)

On the other hand, in Figure 4 shows an AUC score of 0.92 for the 'Excellent' class, 0.81 for the 'Very Good' class, and 0.90 for the 'Good' class. An AUC score close to 1.0 indicates that the model has high performance in distinguishing between universities in Indonesia with 'Excellent', 'Very Good', and 'Good' accreditations.

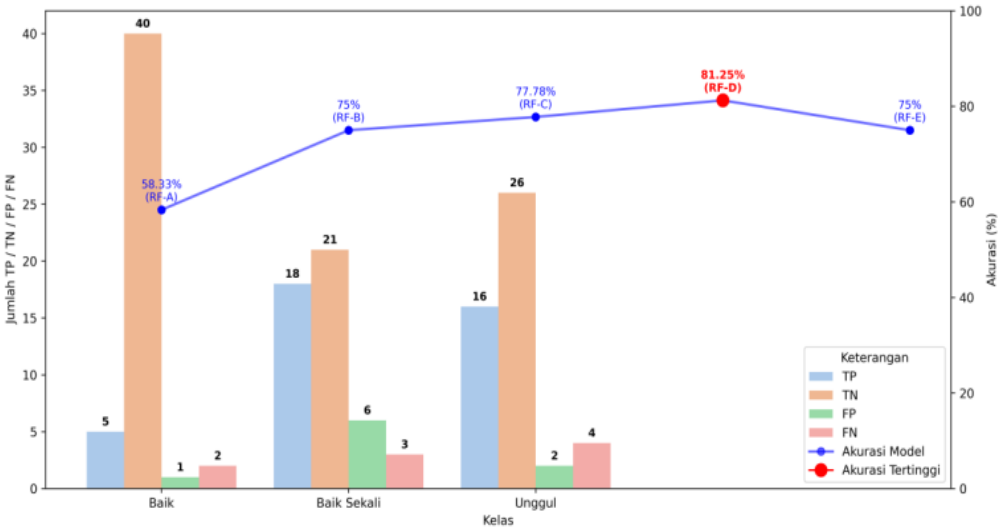


Figure 5. Comparison of the Performance of the Best Random Forest Models

The findings of this study indicate that the Random Forest (RF) model, particularly in the RF-D scenario, has the most optimal performance in measuring the quality of HEIs, both in terms of accuracy, classification balance, and AUC values. These results are in line with research by Mohan et al, which states that Random Forest excels in handling multicategory classification data due to its ability to manage non-linear relationships and noise in educational data (Mohan et al., 2024). Research by Bhuvaneswari & Karthikeyan, also supports these findings, where Random Forest is proven to be superior to Logistic

Regression in the task of predicting academic quality because it provides better accuracy and generalization (Bhuvaneswari & Karthikeyan, 2023).

CONCLUSIONS

This study explores the performance of data mining algorithms using the Random Forest model in effectively and optimally measuring the quality of HEIs in Indonesia. Based on a series of experiments on five RF model scenarios (RF-A to RF-E) with different training and test data ratios, the RF-D model (60:40 ratio) performed best with an accuracy of 81.25%, outperforming the other models in terms of classification balance and prediction accuracy. The RF-D model evaluation results indicate that it is capable of classifying HEIs into three quality categories—Good, Very Good, and Excellent—with relatively high precision, recall, and F1-score values for each class. The Excellent category performed best with 89% precision and 80% recall, while Very Good had the highest sensitivity (86% recall). Furthermore, the RF-D model's AUC score, which approached 1.0 for all categories (Excellent: 0.92, Very Good: 0.81, and Good: 0.90), indicates the model's excellent discriminatory ability in distinguishing higher education quality.

This study also shows that the selection of the training-to-test data ratio significantly impacts model performance. An overly unequal ratio (as in RF-A) risks overfitting, while an overly balanced ratio (as in RF-E) can potentially decrease accuracy due to underfitting. Thus, the use of Random Forest with the appropriate configuration has proven effective in measuring the quality of HEIs in Indonesia.

Although this study successfully demonstrates that the Random Forest model can be used effectively to measure the quality of HEIs in Indonesia, several limitations warrant consideration. First, the data used in this study is relatively limited and does not fully represent all HEIs in Indonesia. This could potentially impact the model's accuracy in learning the patterns and characteristics of each category of higher education quality. Furthermore, the model development process has not fully optimized advanced feature engineering or hyperparameter tuning techniques that could further enhance model performance. Therefore, there is still room for technical improvement in the development of the classification model.

Although this study successfully demonstrates that the Random Forest model can be used effectively to measure the quality of HEIs in Indonesia, several limitations warrant consideration. First, the data used in this study is relatively limited and does not fully represent all HEIs in Indonesia. This could potentially impact the model's accuracy in learning the patterns and characteristics of each category of higher education quality. Furthermore, the model development process has not fully optimized advanced feature engineering or hyperparameter tuning techniques that could further enhance model performance. Therefore, there is still room for technical improvement in the development of the classification model.

ACKNOWLEDGEMENT

The authors would like to express their deepest gratitude to Universitas Islam Negeri Maulana Malik Ibrahim Malang, especially the Master of Informatics Program, for their support and facilities provided during this research and manuscript preparation. Special thanks and appreciation are also extended to the supervisors and examiner lecturers for their valuable guidance, constructive feedback, and insightful suggestions that greatly contributed to the completion of this research.

REFERENCES

Andriani, A. (2012). Penerapan Algoritma C4. 5 Pada Program Klasifikasi Mahasiswa Dropout. *Seminar Nasional Matematika*, 139–147.

Angkotasan, S., & Watianan, S. (2021). Faktor-faktor yang mempengaruhi peningkatan mutu pendidikan di kampus stia alazka ambon. *KOMUNITAS: Jurnal Ilmu Sosiologi*, *4*(2), 42–50.

Asmawi, M. R. (2005). Strategi meningkatkan lulusan bermutu di perguruan tinggi. *Makara Human Behavior Studies in Asia*, *9*(2), 66–71.

Bettis, R. A., Helfat, C. E., & Shaver, J. M. (2016). The necessity, logic, and forms of replication. *Strategic Management Journal*, *37*(11), 2193–2203.

Bhuvaneswari, S., & Karthikeyan, S. (2023). A Comprehensive Research of Breast Cancer Detection Using Machine Learning, Clustering and Optimization Techniques. *2023 International Conference on Data Science and Network Security (ICDSNS)*, 1–6.

Breiman, L. (2001). Random Forest. *Kluwer Academic Publishers*, *45*(Machine Learning), 5–32. https://doi.org/10.1023/A:1010933404324

Estrada-Real, A. C., & Cantu-Ortiz, F. J. (2022). A data analytics approach for university competitiveness: the QS world university rankings. *International Journal on Interactive Design and Manufacturing (IJIDeM)*, *16*(3), 871–891. https://doi.org/10.1007/s12008-022-00966-2

Fitrah, M. (2018). Urgensi sistem penjaminan mutu internal terhadap peningkatan mutu perguruan tinggi. *Jurnal Penjaminan Mutu*, *4*(1), 76–86.

Han, J., Kamber, M., & Pei, J. (2011). Data transformation and data discretization. *Data Mining: Concepts and Techniques*, 111–118.

Hernández-Lobato, J. M., Houlsby, N., & Ghahramani, Z. (2014). Probabilistic matrix factorization with non-random missing data. *International Conference on Machine Learning*, 1512–1520.

Undang-Undang Republik Indonesia Nomor 20 Tahun 2003 tentang Sistem Pendidikan Nasional. (2003). Lembaran Negara Republik Indonesia Tahun 2003 Nomor 78. Jakarta: Pemerintah Republik Indonesia.

Peraturan Pemerintah Republik Indonesia Nomor 60 Tahun 1999 tentang Pendidikan Tinggi. (1999). Lembaran Negara Republik Indonesia Tahun 1999 Nomor 115. Jakarta: Pemerintah Republik Indonesia.

Undang-Undang Republik Indonesia Nomor 12 Tahun 2012 tentang Pendidikan Tinggi. (2012). Lembaran Negara Republik Indonesia Tahun 2012 Nomor 158. Jakarta: Pemerintah Republik Indonesia.

Karim, B. A. (2020). Pendidikan Perguruan Tinggi Era 4.0 Dalam Pandemi Covid-19 (Refleksi Sosiologis). *Education and Learning Journal*, *1*(2), 102–112.

Mitra, M. (2023). Urgensi Percepatan Penjaminan Mutu Perguruan Tinggi Melalui Akreditasi Perguruan Tinggi. *Promis*, *4*(2), 91–99.

Mohan, H. G., Kumar, J., & Nandish, M. (2024). A Hybrid Feature Engineering Mechanism Based on Random Forest for Imbalanced Botnet Dataset. *2024 Fourth International Conference on Multimedia Processing, Communication & Information Technology (MPCIT)*, 140–145.

Muschelli III, J. (2020). ROC and AUC with a binary predictor: a potentially misleading metric. *Journal of Classification*, *37*(3), 696–708.

Nasution, M. K. M. (2016). Karya ilmiah dosen & mahasiswa. *Harian Waspada*.

Nulhaqim, S. A., Heryadi, D. H., Pancasilawan, R., & Ferdryansyah, M. (2016). Peranan perguruan tinggi dalam meningkatkan kualitas pendidikan di Indonesia untuk menghadapi Asean community 2015 studi kasus: Universitas Indonesia, Universitas Padjadjaran, Institut Teknologi Bandung. *Share: Social Work Journal*, *6*(2), 197.

Pramana, C., Chamidah, D., Suyatno, S., & Renadi, F. (2021). Strategies to Improved Education Quality in Indonesia: A Review. *Turkish Online Journal of Qualitative Inquiry*, *12*(3).

Pramana, R., Suherman, A., & Setiawan, D. (2021). Evaluasi standar pendidikan tinggi dan implementasinya di Indonesia. Jurnal Pendidikan dan Kebijakan, 15(2), 112–124.

Kemendikbudristek. (2022). Statistik Pendidikan Tinggi 2022. Available at: https://pddikti.kemdikbud.go.id

Rohmawati, N., Defiyanti, S., & Jajuli, M. (2015). Implementasi Algoritma K-Means Dalam Pengklasteran Mahasiswa Pelamar Beasiswa. *Jurnal Ilmiah Teknologi Infomasi Terapan*, *1*(2).

Romero, C., & Ventura, S. (2007). Educational data mining: A survey from 1995 to 2005. *Expert Systems with*

Applications, *33*(1), 135–146.

Singgih, M. L. (2008). Faktor-faktor yang mempengaruhi kualitas pendidikan pada Perguruan Tinggi. *Prosiding Seminar Nasional Teknoin*.

Sinsomboonthong, S. (2022). Performance comparison of new adjusted min-max with decimal scaling and statistical column normalization methods for artificial neural network classification. *International Journal of Mathematics and Mathematical Sciences*, *2022*(1), 3584406.

Syahrir, N., Ansari, M. I., & Basir, I. (2023). Pendampingan tata kelola keuangan dan pemasaran berbasis digital dalam meningkatkan penjualan. *Jurnal Abdi Insani*, *10*(3), 1639–1646.

Ying, X. (2019). An overview of overfitting and its solutions. *Journal of Physics: Conference Series*, *1168*, 22022.

Sampe, N., & Arifin, Z. (2024). Internal Quality Assurance System in Indonesia Higher Education: Literature Review. *Indonesian Journal of Educational Research and Review*, 7(1), 73–84.

Utami, S. G. A. (2025). Pengembangan model evaluasi universitas berbasis machine learning. *Jurnal Teknologi Pendidikan*, 15(2), 120–132.