Journal of Information Systems and Informatics
Vol. 7, No. 1, March 2025 e-ISSN: 2656-4882 p-ISSN: 2656-5935
DOI: 10.51519/journalisi.v7i1.977 Published By DRPM-UBD

Predicting Student Loyalty in Higher Education Using Machine Learning: A Random Forest Approach

Qoriani Widayati[1], Kusworo Adi[2], R Rizal Isnanto[3], Eka Puji Agustini[4], Dewa Rizki Rahmat Julianto[5], Fawwaz Bimo Prakasa[6]

[1,2,3] Doctoral Program of Information Systems, School of Postgraduate Studies, Diponegoro University, Semarang, Indonesia
[1]Accounting Information System, Bina Darma University, Palembang, Indonesia
[4,5,6]Information System Department, Bina Darma University, Palembang, Indonesia
Email: [1]qoriani_widayati@binadarma.ac.id [2]eka_puji@binadarma.ac.id

Abstract

Student loyalty is a crucial factor supporting the sustainability of higher education institutions. The aim of this study is to predict student loyalty using a machine learning approach, specifically the random forest algorithm. The data for this research were collected through a questionnaire that included variables such as service quality, emotional attachment, brand satisfaction, brand trust, and socio-economic conditions, distributed to 107 students in Palembang. The resulting dataset was processed through preprocessing, model training, and performance evaluation, employing metrics such as accuracy, precision, recall, and F1-score. The analysis using the random forest algorithm achieved an accuracy of 90.9%. These findings are expected to provide valuable insights for higher education institutions in developing more effective strategies to enhance student loyalty.

Keywords: student loyalty, random forest, machine learning

1. INTRODUCTION

Student loyalty to higher education institutions is crucial for the success and growth of these institutions. Building positive relationships with students and providing satisfying learning experience are long-term investments that higher education institutions can make. These efforts not only contribute to the long-term success of the institution but also foster increased student loyalty [1]. There are several reasons why student loyalty is crucial for higher education institutions. Firstly, it plays a key role in student retention, enhances the institution's image and reputation—particularly regarding institutional accreditation, encourages student participation and engagement in campus activities, and has a positive impact on the teaching system. Moreover, loyal students contribute to the institution's financial stability and strengthen alumni relations, as they remain connected to the institution and continue to contribute even after graduation [2].

This work is licensed under a Creative Commons Attribution 4.0 International License.

p-ISSN: **2656-5935** http://journal-isi.org/index.php/isi e-ISSN: **2656-4882**

In the context of higher education, a loyal student is one who remains enrolled at the institution and does not transfer to another university until completing their studies. Additionally, loyal students are more likely to pursue further education at the same institution and to provide positive word-of-mouth recommendations, sharing their experiences with family, friends, and acquaintances whenever the opportunity arises. Therefore, it is essential for policymakers and administrators in higher education institutions to identify the factors that contribute to student loyalty [3]. Higher education institutions encounter several challenges in enhancing student loyalty, such as limitations in managing student and alumni data—both in terms of human resources and technology—along with issues related to the quality of services offered, internal factors affecting students, and the development strategies implemented by the institution. Therefore, it is crucial for higher education institutions to devise strategies for effectively managing student loyalty. This loyalty is most apparent when students are satisfied with the quality of services provided by the institution, including tangibility, reliability, responsiveness, assurance, and empathy [4].

Several previous studies have employed Structural Equation Modeling (SEM) to identify the factors that influence student loyalty to higher education institutions. The study conducted by [4] used SEM to examine the relationship between service quality and student loyalty, while [5] assessed the connection between the quality of student-teacher relationships and student loyalty. However, this SEM approach has limitations in accurately predicting student behavior.

The rapid advancement of technology has facilitated the integration of machine learning algorithms, particularly in the field of education. Numerous studies have explored their applications in various educational contexts. For instance, machine learning is widely utilized in recommendation systems designed to personalize learning materials and educational content based on individual student profiles, including their educational background, interests, and skills [6]. Moreover, machine learning serves as a framework for predicting academic pathways and optimizing academic planning through the Apriori algorithm [7]. Additionally, in the education sector, machine learning is employed for meta-analysis, enabling the development of innovative learning techniques that enhance efficiency, precision, and research quality [8].

The machine learning algorithm used in this study is the random forest algorithm, a classification algorithm that can be applied to determine the most important criteria associated with a specific research object. Below are studies on higher education institutions that have utilized the random forest algorithm. The study by [9] employed the random forest algorithm to determine the best classification of the factors influencing final course grades. In this study, the random forest algorithm achieved an evaluation result of 90.33%.

p-ISSN: **2656-5935** http://journal-isi.org/index.php/isi e-ISSN: **2656-4882**

The study by [10] used the random forest algorithm to identify the attributes influencing students' academic performance and to develop a valid model for predicting performance based on their high school GPA. The random forest algorithm in this study achieved an evaluation result of 91.32%. The Random Forest algorithm has been proficiently employed across diverse educational settings to forecast student performance, exhibiting a higher degree of accuracy in comparison to conventional methodologies such as linear regression. For example, research concerning veterinary students demonstrated that Random Forest models attained an exceptional accuracy range of 96.1% to 99% in predicting scholarly achievement, underscoring the importance of both academic and financial variables over demographic considerations [11]. Conversely, a separate investigation revealed that although Random Forest delivered consistent predictions for student admissions, linear regression produced fewer overall errors, indicating its dependability in certain forecasting scenarios [12].

Furthermore, Random Forest has been fine-tuned for the prediction of student dropout rates, illustrating its proficiency in identifying students at risk through refined parameter optimization, which is vital for the implementation of timely interventions [13]. Collectively, these investigations underscore the adaptability and efficacy of Random Forest within the domain of educational analytics, particularly in the realms of predicting academic performance and retention rate[14] [15]. While research utilizing the random forest algorithm to predict student loyalty in higher education is still limited, this study aims to fill that gap by applying the random forest algorithm to predict student loyalty based on survey data, which includes variables such as service quality, emotional attachment, brand satisfaction, brand trust, and socio-economic conditions. This study is expected to contribute to the academic community by providing valuable insights for similar research and offering useful perspectives for higher education administrators in developing strategies for institutional management.

2. METHODS

This study employs a quantitative approach, focusing on the application of the random forest algorithm to analyze survey data from respondents. The research stages are detailed in Figure 1.

2.1 Dataset

The data for this study were collected through an online survey distributed via Google Forms to students in Palembang. The questionnaire covered several variables, including service quality, emotional attachment, brand satisfaction, brand trust, socio-economic conditions, and student loyalty. A total of 107 student respondents participated in completing the survey. Table 1 presents a detailed

p-ISSN: **2656-5935** http://journal-isi.org/index.php/isi e-ISSN: **2656-4882**

breakdown of the variables or attributes in the dataset used.

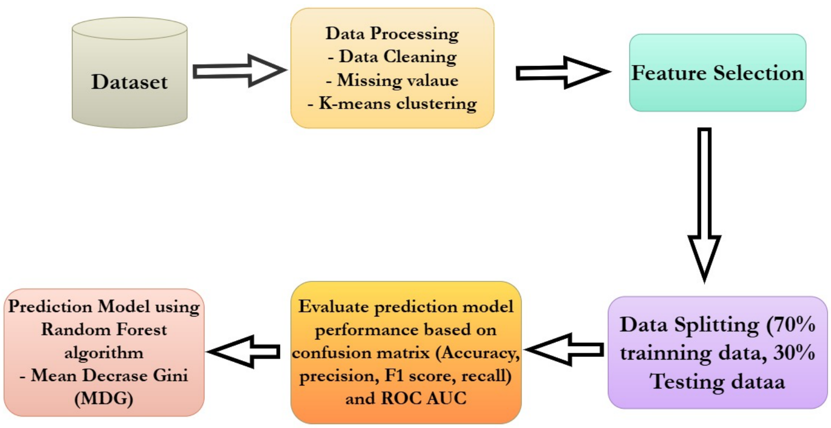

Figure 1. Research diagram

Table 1. Atribute and description

Name Atribut	Information
Service_quality	In what ways do universities offer both academic and non-academic services? [4]
Emotional_attachment	What factors contribute to the development of emotional attachment between students and universities?[16]
Brand_satisfaction	What is the current level of student satisfaction with higher education institutions, and what factors influence this perception?[17]
Brand_Trust	To what extent do students feel confident that their college will take the necessary steps to support them in achieving their academic goals?[16]
Economic_social	In what ways might students' economic and social circumstances shape their behavior, influence their perceptions, or affect their commitment to college?[18]
Student_loyalty	Student loyalty encompasses the degree of connection students feel toward an institution, as reflected in their attitudes and behaviors that demonstrate this bond.[5]

2.2 Preprocessing

At this stage, data cleaning is performed, where the initial dataset obtained from the input process is further processed using machine learning techniques. This begins with data preprocessing to address missing values, remove noise, and correct inconsistencies through data cleaning and transformation. The objective at

p-ISSN: **2656-5935** http://journal-isi.org/index.php/isi e-ISSN: **2656-4882**

this stage is to ensure the dataset is ready and aligned with the research requirements.

2.3 Clustering with K-means

In this study, student loyalty categories are grouped using the K-Means clustering method before being classified with the Random Forest algorithm. The selection of K-Means is based on its ability to identify hidden patterns in data without requiring initial assumptions about data distribution [19]. In this study, students are categorized into four loyalty groups: not loyal, temporary loyal, hidden loyal, and premium loyal (Table 2). This clustering process is conducted to facilitate a better understanding of student segmentation before proceeding with further analysis using the Random Forest algorithm.

Table 2. loyalty cluster

Categories of loyalty	Cluster
No loyalty	0
Inertia loyalty	1
Latent loyalty	2
Premium loyalty	3

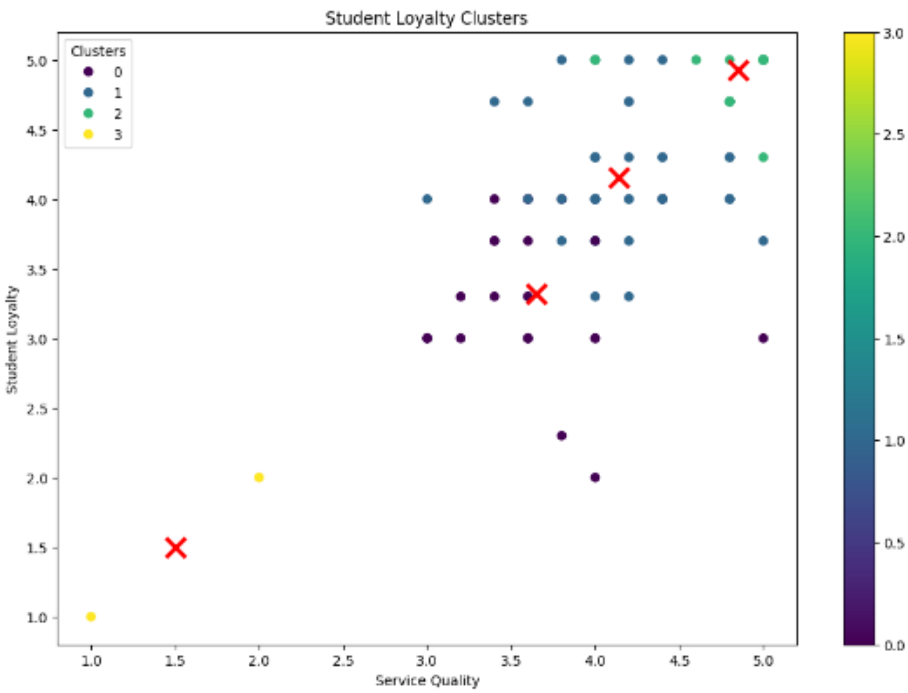

Figure 2. Clustering of student loyalty

p-ISSN: **2656-5935** http://journal-isi.org/index.php/isi e-ISSN: **2656-4882**

Figure 2 illustrates the visualization of student loyalty clustering. Following this, the characteristics of the variables are derived based on the averages and the number of students in each category, as presented in Table 3.

Table 3. Characteristics of the variables

Categories loyalty	**Service_quality**	**Brand_satisfaction**	**Emotional_attachment**	**Brand_trust**	**Economic_social**	**Student_loyalty**	**Total_Student**
Premium loyalty	4,09	4,01	3,71	3,91	4,08	4,12	34
Inertia loyalty	3,66	3,56	3,18	3,34	3,52	3,29	24
Latent loyalty	1,5	1,5	1,5	1,5	1,5	1,5	2
No loyalty	4,82	4,85	4,43	4,86	4,82	4,88	47

Table 3 explains that premium loyalty group consists of individuals who score highly across all dimensions, with an average rating above 4.0. This group exhibits exceptional levels of satisfaction and loyalty. The Inertia loyalty group is characterized by moderate scores across all aspects, with an average ranging from 3.2 to 3.6. This group demonstrates a satisfactory level of loyalty, although there is still room for improvement. The Latent loyalty group represents the smallest cluster, with a low average score (1.5) across all aspects. Their loyalty may stem from factors that are not measured in this study. No loyalty group has the highest scores across all aspects, with an average above 4.8. This may indicate that they have very high expectations. The next step involves splitting the data into 70% training data and 30% testing data.

2.4 Machine Learning method using Random Forest Algorithm

At this stage, training and validation are conducted to select the model. Subsequently, the importance of each variable is calculated using Mean Decrease Gini (MDG), where a higher MDG value indicates a greater influence of the corresponding independent variable. The random forest algorithm is essential for determining the value of the variable mmm, the number of predictor variables

p-ISSN: **2656-5935** http://journal-isi.org/index.php/isi e-ISSN: **2656-4882**

randomly selected, and for processing the values of kkk trees to achieve optimal results. The recommended value of kkk is used in the bagging method. Bagging, which stands for bootstrap aggregating, involves taking different bootstrap samples, L(θ)L(\theta)L(θ), of size nnn from the training set LLL of size NNN, with each sample serving as a modified learning set for the construction of a new tree. Each predictor tree, TL(θ)T_L(\theta)TL(θ), depends on a random vector θ\thetaθ, which represents a sample drawn from the set LLL. The prediction result is determined by the majority vote or the average of all decision trees, as represented by the following formula as shown in Equation 1 [20]:

$$y^2=\{T_{L(\theta)}\}_1^k \tag{1}$$

The sample size of the explanatory variable mmm when using the random forest method significantly influences the correlation and strength of each tree. To determine the value of mmm, the number of predictor variables is randomly selected, with ppp representing the total number of independent variables. This can be expressed as follows [20].

For the classification process, the value of mmm is determined using the formula $|\sqrt{p}|$, where the value of the smallest node or terminal node is 1. For the regression process, the value of mmm is determined using the formula $\left|\frac{p}{3}\right|$ where the value of the smallest node or terminal node is 5. To determine the value of mmm by observing the out-of-bag (OOB) error, there are three methods, as shown in the equation 2 to 4.

$$m=\frac{1}{2}|p| \tag{2}$$

$$m=\sqrt{p} \quad 3)$$

$$m=2\times\sqrt{p} \tag{4}$$

The value represents the total number of variables. Proper selection of mmm will result in a random forest where the correlation between trees is sufficiently low, yet the strength of each tree is high, as indicated by a small out-of-bag (OOB) error. The OOB error depends on both the correlation between trees and the strength of each individual tree in the random forest process. Specifically, an increase in correlation tends to raise the OOB error, while an increase in the number of trees generally reduces the OOB error. The out-of-bag (OOB) error is calculated by comparing the classification results, which are the predictions made by the random forest algorithm [20]. In this study, the Random Forest algorithm was selected due to its ability to generate reliable predictions and identify the most influential features or variables in the analysis. The key parameters of this model

p-ISSN: **2656-5935** http://journal-isi.org/index.php/isi e-ISSN: **2656-4882**

are carefully selected to balance accuracy and computational efficiency. The number of trees (*n_estimators*) is set to 100 to achieve an optimal trade-off between performance and processing time. The maximum depth (*max_depth*) is limited to 10 to prevent excessive complexity and reduce the risk of overfitting. To ensure model stability across different data subsets, a 10-fold cross-validation method is employed. Additionally, feature selection is performed using Mean Decrease in Gini (MDG), which provides insights into the most influential variables in predicting student loyalty.

The model evaluation is conducted using several metrics, including accuracy, precision, recall, and F1-score. Additionally, the Receiver Operating Characteristic (ROC) curve and the Area Under the Curve (AUC) are utilized to assess the model's ability to distinguish between different loyalty categories. All analyses are performed using Python, leveraging libraries such as Scikit-learn.

3. RESULTS AND DISCUSSION

3.1 Prediction Model Using Machine Learning Algorithm

The prediction results using the Random Forest algorithm demonstrate excellent performance in classifying student loyalty, achieving an accuracy of 82%. This study also compares the performance of Random Forest, Decision Tree, and Gradient Boosting algorithms to evaluate the effectiveness of the Random Forest approach. Table 4 presents a comparative evaluation of the metrics for the three algorithms used

Table 4. Performance Comparison of Machine Learning Algorithms in Predicting Student Loyalty.

Algorithm	Accuracy (%)	Precision	Recall	F1 Score
Random Forest	82	0,88	0,88	0,88
Decision tree	73	0,8	0,88	0,82
Gradient boosting	79	0,84	0,88	0,86

The evaluation results highlight performance differences among the Random Forest, Decision Tree, and Gradient Boosting algorithms. The Random Forest algorithm achieves the highest accuracy at 82%, followed by Gradient Boosting at 79% and Decision Tree at 73%. Random Forest demonstrates strong generalization capabilities, maintaining a well-balanced trade-off between recall and precision, both of which approach a score of 0.88. In addition to Random Forest, Gradient Boosting also delivers competitive performance, with precision, recall, and F1-score values closely matching those of the Random Forest model. One of the key advantages of Gradient Boosting is its ability to handle complex data while producing balanced predictions. Figure 3 provides a visual representation of the performance of all three algorithms.

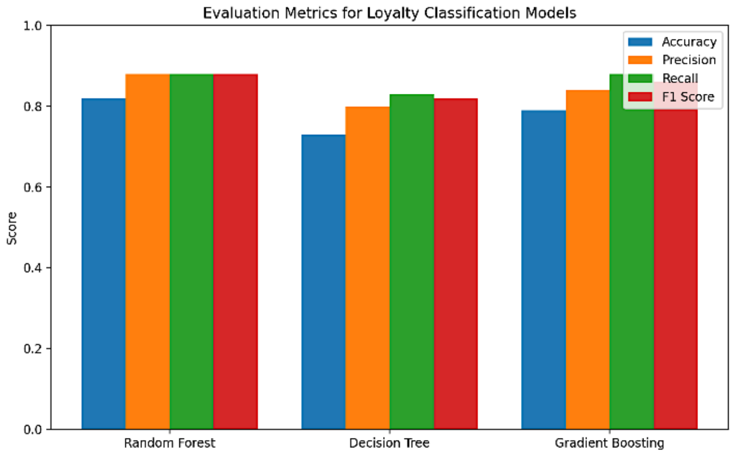

Figure 3. Accuracy Comparison of Machine Learning Algorithms.

This study also evaluates the ability of the algorithms to distinguish between student loyalty categories using the Receiver Operating Characteristic (ROC) curve and the Area Under the Curve (AUC). The ROC curve illustrates the relationship between the True Positive Rate (TPR) and the False Positive Rate (FPR) in the predictive model, while the AUC quantifies the model's overall effectiveness in accurately classifying data. Figure 4 presents the ROC curves for the three algorithms used in this study.

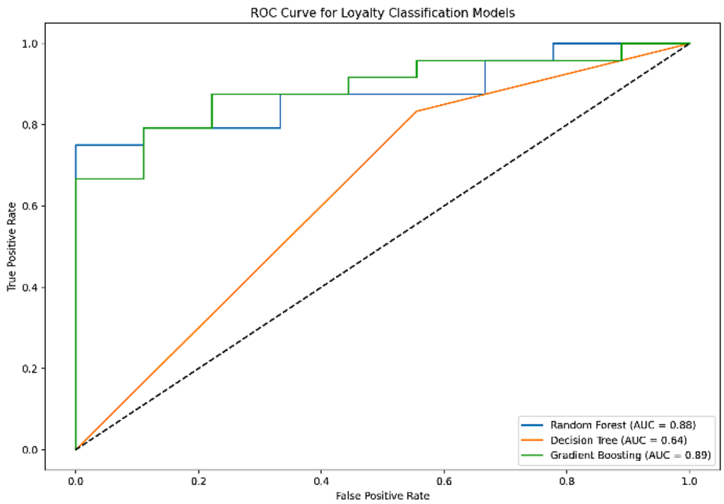

Figure 4. ROC Curves for Decision Tree, Random Forest, and Gradient Boosting Models

p-ISSN: **2656-5935** http://journal-isi.org/index.php/isi e-ISSN: **2656-4882**

The ROC curve illustrates the trade-off between the true positive rate and the false positive rate for each model, providing insight into their classification performance. The results indicate that both Random Forest and Gradient Boosting exhibit greater reliability in classifying student loyalty.

3.2 Analysis of Factors Influencing Student Loyalty

The Mean Decrease in Gini (MDG), reflects the importance of each variable in the Random Forest model. The higher the MDG value, the greater the influence of the corresponding independent variable [21].

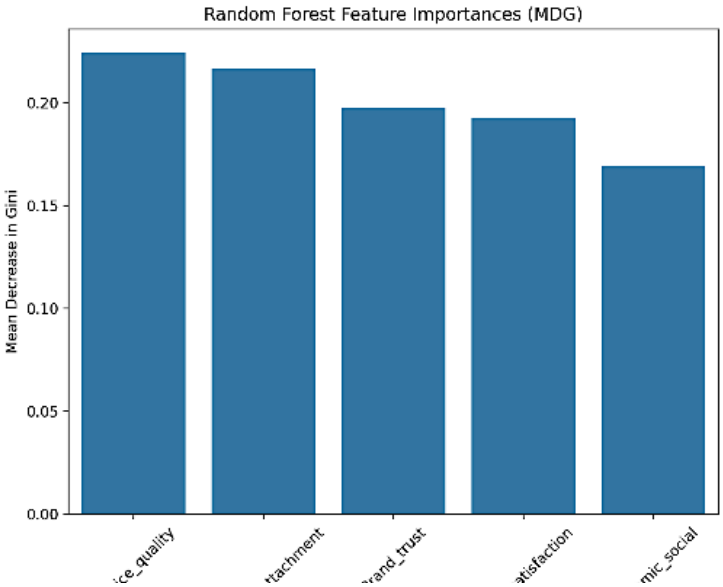

Figure 5. Feature Importance Analysis in the Random Forest Model

The analysis results presented in Figure 5 indicate the following Mean Decrease in Gini (MDG) values: service quality (0.105), brand satisfaction (0.252), emotional attachment (0.269), brand trust (0.185), and economic-social factors (0.188). These findings reveal that brand satisfaction and emotional attachment are the most dominant factors, as they have the highest MDG values. This suggests that significant improvements in brand satisfaction and emotional attachment can effectively reduce uncertainty in the model. Efforts to enhance brand satisfaction and emotional attachment have the potential to significantly impact loyalty, particularly through improvements in service quality and brand identity as integral components of a comprehensive marketing strategy. This aligns with the findings of [22], which emphasize the importance of a well-rounded marketing approach that prioritizes brand satisfaction and emotional connection to strengthen customer loyalty across various sectors.

p-ISSN: 2656-5935 http://journal-isi.org/index.php/isi e-ISSN: 2656-4882

3.3 Discussion

Machine learning algorithms can effectively predict student loyalty with higher accuracy compared to traditional statistical methods such as Structural Equation Modeling (SEM). In particular, Random Forest and Gradient Boosting excel at capturing nonlinear patterns and complex relationships between variables. Table 5 presents a comparison of accuracy levels and interpretability between statistical models and machine learning-based approaches.

Table 5. Comparison of SEM and Machine Learning

Algorithm	Accuracy (%)	Linier Relationship	Interpretability
SEM	7	Strong	High
Random Forest	82	Weak/non linier	Moderate
Decision tree	73	Weak/non linier	low
Gradient boosting	79	Moderat	High

The analysis results indicate that Gradient Boosting achieves the highest predictive accuracy (0.85), followed by Random Forest (0.82), Decision Tree (0.78), and Structural Equation Modeling (SEM) (0.70). Machine learning models, such as Gradient Boosting and Random Forest, excel at capturing complex nonlinear patterns in data. However, Gradient Boosting's lower interpretability poses a challenge when explaining linear relationships, whereas SEM offers stronger interpretability. Among these methods, Decision Tree emerges as a strong candidate for higher education data analysis due to its balance between relatively high accuracy and good interpretability. This comparison highlights that integrating machine learning techniques with traditional statistical methods can provide deeper insights and support more data-driven decision-making. Moreover, these findings align with the study by [23], which concluded that machine learning algorithms outperform linear regression and SEM in predicting customer behavior and student loyalty. Furthermore, the SEM analysis indicates that the relationship between brand trust and student loyalty is not statistically significant. However, this finding contrasts with the results obtained using the Random Forest algorithm, which identifies brand trust as a key variable with a substantial impact on student loyalty. Table 6 presents a comparison of SEM and Random Forest in predicting student loyalty, highlighting their respective performance differences.

Table 6. Comparison of SEM and Random Forest Performance

Variabel	SEM (coefficient,nsignificance)	Random forest (MDG)
Service quality	0,134 (Significant)	0,105

p-ISSN: **2656-5935** http://journal-isi.org/index.php/isi e-ISSN: **2656-4882**

Variabel	SEM (coefficient,nsignificance)	Random forest (MDG)
Brand satisfaction	0,278 (Significant)	0,252
Emotional attachment	0,315 (Significant)	0,269
Brand trust	0,092 (not Significant)	0,185
Economic & Social	0,127 (Significant)	0,188

This study has significant practical implications for higher education, particularly in data-driven decision-making. With high predictive accuracy and the ability to identify key variables such as brand satisfaction and emotional attachment as crucial indicators, universities can implement more targeted marketing strategies to enhance student loyalty. This aligns with previous research by [24], which found that brand satisfaction—defined as students' overall satisfaction with their university—plays a critical role in influencing student loyalty in Nigerian higher education institutions. Furthermore, universities must prioritize ethical considerations and student data privacy by ensuring compliance with data protection regulations and implementing transparent policies regarding data usage.

4. CONCLUSION

The Random Forest algorithm has proven to be highly effective in predicting student loyalty in higher education. With an accuracy rate of 82%, this model outperforms conventional statistical approaches such as SEM in classifying student loyalty. Moreover, the feature importance analysis reveals that brand satisfaction and emotional attachment are the two most influential factors contributing significantly to student loyalty. The findings of this study further emphasize that fostering student satisfaction and engagement with the institution—both through academic and non-academic services—is a key strategy for improving student retention. The application of machine learning in managing student loyalty provides valuable insights for higher education institutions, enabling them to identify student segments with low loyalty levels and implement targeted strategies to enhance retention based on data-driven analysis. However, despite its contributions to leveraging machine learning for student loyalty prediction, this study has certain limitations. One major constraint is the relatively small sample size of 107 students, all from a single geographic region, which limits the generalizability of the findings. Future research should consider expanding the dataset and involving a more diverse population from multiple institutions across different geographic areas to improve the robustness of the results. Additionally, future studies could incorporate additional factors, such as students' academic performance, institutional reward systems, and extracurricular activities, to further refine the model's predictive capabilities.

p-ISSN: **2656-5935** http://journal-isi.org/index.php/isi e-ISSN: **2656-4882**

REFERENCES

[1] I. Snijders, L. Wijnia, R. M. J. P. Rikers, and S. M. M. Loyens, 'Alumni loyalty drivers in higher education', *Soc. Psychol. Educ.*, vol. 22, no. 3, pp. 607–627, 2019, doi: 10.1007/s11218-019-09488-4.

[2] S. Todea, A. A. Davidescu, N. A. Pop, and T. Stamule, 'Determinants of Student Loyalty in Higher Education: A Structural Equation Approach for the Bucharest University of Economic Studies, Romania', *Int. J. Environ. Res. Public Health*, vol. 19, no. 9, 2022, doi: 10.3390/ijerph19095527.

[3] L. J. Wong, P. S. Ling, and T. H. Y. Ling, 'A conceptual framework for higher education student loyalty from the green marketing perspective', *High. Educ. Ski. Work. Learn.*, vol. 13, no. 2, pp. 387–402, 2023, doi: 10.1108/HESWBL-08-2022-0165.

[4] T. T. Borishade, O. O. Ogunnaike, O. Salau, B. D. Motilewa, and J. I. Dirisu, 'Assessing the relationship among service quality, student satisfaction and loyalty: the NIGERIAN higher education experience', *Heliyon*, vol. 7, no. 7, p. e07590, 2021, doi: https://doi.org/10.1016/j.heliyon.2021.e07590.

[5] I. Snijders, L. Wijnia, R. M. J. P. Rikers, and S. M. M. Loyens, 'Building bridges in higher education: Student-faculty relationship quality, student engagement, and student loyalty', *Int. J. Educ. Res.*, vol. 100, Jan. 2020, doi: 10.1016/j.ijer.2020.101538.

[6] S. I. Q. I. LI and D. LI, 'Research on Personalized Learning Recommendation System Based on Machine Learning Algorithm', *Scalable Comput.*, vol. 26, no. 1, pp. 432–440, 2025, doi: 10.12694/scpe.v26i1.3844.

[7] M. Muhairat, W. Alzyadat, A. Shaheen, A. Alhroob, and A. Nasser Asfour, 'Leveraging Machine Learning for Predictive Pathways in Higher Education: A Case Study at Al-Zaytoonah University of Jordan', *SSRG Int. J. Electron. Commun. Eng.*, vol. 11, no. 11, pp. 28–44, 2024, doi: 10.14445/23488549/IJECE-V11I11P104.

[8] O. Chernikova, M. Stadler, I. Melev, and F. Fischer, 'Using machine learning for continuous updating of meta-analysis in educational context', *Comput. Human Behav.*, vol. 156, no. December 2023, p. 108215, 2024, doi: 10.1016/j.chb.2024.108215.

[9] M. Nachouki, E. A. Mohamed, R. Mehdi, and M. Abou Naaj, 'Student course grade prediction using the random forest algorithm: Analysis of predictors' importance', *Trends Neurosci. Educ.*, vol. 33, p. 100214, 2023, doi: 10.1016/j.tine.2023.100214.

[10] M. Nachouki and M. A. Naaj, 'Predicting Student Performance to Improve Academic Advising Using the Random Forest Algorithm', *Int. J. Distance Educ. Technol.*, vol. 20, no. 1, pp. 1–17, 2022, doi: 10.4018/IJDET.296702.

p-ISSN: **2656-5935** http://journal-isi.org/index.php/isi e-ISSN: **2656-4882**

[11] A. E. Hooper SE, Ragland N, 'Random forest models reveal academic and financial factors outweigh demographics in predicting completion of a year-round veterinary program', *J Am Vet Med Assoc*, vol. 263, no. (2), pp. 1–9, doi: 10.2460/javma.24.08.0501.
[12] M. Nurdin, 'Analysis Forecasting Students Using Random Forest and Linear Regression Algorithms', vol. 8, no. 4, pp. 2369–2376, 2024.
[13] D. Kumar, A. Kothiyal, R. Kumar, C. Hemantha, and R. Maranan, 'Random Forest approach optimized by the Grid Search process for predicting the dropout students', *2024 Int. Conf. Innov. Challenges Emerg. Technol. ICICET 2024*, pp. 1–6, 2024, doi: 10.1109/ICICET59348.2024.10616372.
[14] G. Petrea, R. A. Puiu, B. C. Mocanu, and O. M. K. Al-Dulaimi, 'Determining the Degree of Conviction of Students in University Selection Using the Random Forest Algorithm: An Approach for Adaptive and Personalized Decision Support System in Education', *Proc. - RoEduNet IEEE Int. Conf.*, pp. 1–6, 2024, doi: 10.1109/RoEduNet64292.2024.10722377.
[15] Y. Miao and Y. Xu, 'Random Forest-Based Analysis of Variability in Feature Impacts', *2024 IEEE 2nd Int. Conf. Image Process. Comput. Appl. ICIPCA 2024*, pp. 1130–1135, 2024, doi: 10.1109/ICIPCA61593.2024.10708791.
[16] S. Todea, A. A. Davidescu, N. A. Pop, and T. Stamule, 'Determinants of Student Loyalty in Higher Education: A Structural Equation Approach for the Bucharest University of Economic Studies, Romania', *Int. J. Environ. Res. Public Health*, vol. 19, no. 9, 2022, doi: 10.3390/ijerph19095527.
[17] S. Huston, E. Huston, and M. Kozlowski, 'Learning dispositif and emotional attachment: A preliminary international analysis', *Educ. Sci.*, vol. 9, no. 4, pp. 1–22, 2019, doi: 10.3390/educsci9040279.
[18] S. Aghaei, Y. Shahbazi, M. Pirbabaei, and H. Beyti, 'A hybrid SEM-neural network method for modeling the academic satisfaction factors of architecture students', *Comput. Educ. Artif. Intell.*, vol. 4, no. June 2022, p. 100122, 2023, doi: 10.1016/j.caeai.2023.100122.
[19] M. Wati, W. H. Rahmah, N. Novirasari, Haviluddin, E. Budiman, and Islamiyah, 'Analysis K-Means Clustering to Predicting Student Graduation', *J. Phys. Conf. Ser.*, vol. 1844, no. 1, 2021, doi: 10.1088/1742-6596/1844/1/012028.
[20] L. Breiman, 'Random Forest', *Mach. Learn.*, vol. 45, pp. 5–32, 2001, doi: 10.1007/978-3-030-62008-0_35.
[21] Hong Han; Xiaoling Guo; Hua Yu, 'Variable selection using Mean Decrease Accuracy and Mean Decrease Gini based on Random Forest', in *2016 7th IEEE International Conference on Software Engineering and Service Science (ICSESS)*, China: IEEE, 2016.

p-ISSN: **2656-5935** http://journal-isi.org/index.php/isi e-ISSN: **2656-4882**

[22] H. S. Utomo, H. Sutanto, and R. D. Fadma, 'The Effect of E-Service Quality on Brand Love : E-Satisfaction as a Mediator', vol. 0, no. 01, pp. 558–566, 2025, doi: 10.47191/jefms/v8.
[23] S. K. Rajak, 'Machine Learning Models for Predicting Consumer Behaviour Trends', vol. 19, no. 10, pp. 674–681, 2021, doi: 10.48047/nq.2021.19.10.NQ21218.
[24] T. T. Borishade, O. O. Ogunnaike, O. Salau, B. D. Motilewa, and J. I. Dirisu, 'Assessing the relationship among service quality, student satisfaction and loyalty: the NIGERIAN higher education experience', *Heliyon*, vol. 7, no. 7. Elsevier Ltd, 2021. doi: 10.1016/j.heliyon.2021.e07590.