Pak. j. life soc. Sci. (2024), 22(2): 7844-7856 E-ISSN: 2221-7630;P-ISSN: 1727-4915

Pakistan Journal of Life and Social Sciences
www.pjlss.edu.pk

https://doi.org/10.57239/PJLSS-2024-22.2.00591

RESEARCH ARTICLE

Prediction of University Student Satisfaction in the Educational Experience Using Decision Trees and Logistic Regression: A Comparative Analysis

Milton Antonio López-Cueva[1*], Soledad Jackeline Zegarra-Ugarte[2], Edward Hipolito Torres-Pacheco[3], Rodolfo Fredy Arpasi-Chura[4], Porfirio Enríquez-Salas[5], Peregrino Melitón López-Paz[6]

[1,2,3,5,6] Universidad Nacional del Altiplano, Puno, Peru

[4] Universidad Andina Néstor Cáceres Velásquez, Puno, Peru

ARTICLE INFO

Received: Aug 11, 2024

Accepted: Oct 8, 2024

Keywords

Student satisfaction

Decision trees

Logistic regression

Prediction

Educational quality

***Corresponding Author:**

malopez@unap.edu.pe

ABSTRACT

The objective of this study is to predict student satisfaction in higher education by applying decision tree and logistic regression models. A digital questionnaire was distributed to students at the National University of Altiplano, Peru, resulting in 1,020 valid responses. The dataset included 20 predictor variables, such as teaching quality, access to resources, and academic workload, alongside one dependent variable: student satisfaction. Decision trees and logistic regression models were trained with 80% of the data and tested with 20%. The results showed that the decision tree model achieved an accuracy of 95.6% based on the confusion matrix, while logistic regression reached 92.6%. Both models effectively identified key factors influencing student satisfaction, with the decision tree showing slightly better predictive performance. The study concludes that decision trees provide a more precise model for predicting student satisfaction in comparison to logistic regression. These findings highlight the usefulness of machine learning techniques in educational management for identifying areas needing improvement, ultimately enhancing student experiences and academic outcomes. This study provides a reliable model for university administrators to optimize resource allocation and improve educational practices. The use of machine learning models can lead to data-driven decisions that positively impact student satisfaction and institutional performance.

INTRODUCTION

Student satisfaction refers to the degree to which students feel content with their educational experience, including aspects such as the quality of teaching, the academic environment, institutional support (Pan & Cutumisu, 2024), access to resources, the learning environment (Chowdhury et al., 2024), appropriate teaching styles, fulfillment of learning needs (AL Mutawa & Sruthi, 2023), interaction with professors and peers, and the relevance of course content (Ho et al., 2021). This topic is significant in higher education because it influences academic performance, student retention, overall well-being (Ho et al., 2021; Pan & Cutumisu, 2024), motivation, engagement, the quality of education, and the formation of future professionals (Chowdhury et al., 2024). Moreover, high satisfaction can contribute to better mental health and a more positive university experience (Pan & Cutumisu, 2024).

Student satisfaction impacts the perception of teaching quality and available resources, which can motivate educators to improve their methods and pedagogical approaches (Pan & Cutumisu, 2024). Students satisfied with their learning experience are more likely to participate actively and achieve better academic results (AL Mutawa & Sruthi, 2023; Ho et al., 2021). They are more inclined to continue their studies and complete their programs, reducing dropout rates and enhancing institutional stability (Ho et al., 2021; Pan & Cutumisu, 2024). Satisfaction fosters a sense of belonging and loyalty towards the institution, leading students to complete their programs and recommend the institution to others (Ho et al., 2021). Additionally, high student satisfaction can translate into better evaluations and institutional reputation, attracting more students and potentially increasing funding and resources available to the institution (Ho et al., 2021; Pan & Cutumisu, 2024).

Predicting student satisfaction allows institutions to identify areas for improvement in teaching and services, leading to higher quality education (Pan & Cutumisu, 2024). It enables institutions to recognize less satisfied students and provide them with adequate resources and support to enhance their learning experience (AL Mutawa & Sruthi, 2023). This contributes to creating a more positive and effective learning environment, benefiting both students and educators (AL Mutawa & Sruthi, 2023). Understanding student satisfaction helps universities implement strategies that foster retention and reduce dropout rates, ensuring students complete their studies (Pan & Cutumisu, 2024). It allows for the optimization of teaching methods (AL Mutawa & Sruthi, 2023), which not only benefits students but also strengthens the position and sustainability of the institution in the competitive educational landscape (Ho et al., 2021).

Machine learning algorithms are computational methods that enable machines to learn from data and make predictions; they use statistical and mathematical techniques to identify patterns and relationships within data (Alruwais & Zakariah, 2023). Decision trees and logistic regression are two such algorithms utilized in data analysis and data mining (Abidi et al., 2020; Raj & Renumol, 2023), especially in the context of classification and prediction (Almasri et al., 2022). Decision trees are predictive models that use a tree-like structure to make decisions based on data features; they function by dividing data into subsets through decisions that maximize information gain or minimize impurity (Altun et al., 2022; Elouafi et al., 2024; Munshi et al., 2023). They can handle both categorical and numerical variables without the need for prior transformation (Almasri et al., 2022), are efficient for large datasets allowing quick decisions (Jamil & Belkacem, 2024) and are useful in fields such as education (Jamil & Belkacem, 2024).

Logistic regression is a statistical model (Mazouch et al., 2018; Palacios et al., 2021) used to predict the probability of a binary event (e.g., yes/no, success/failure) based on one or more independent variables or factors (Almasri et al., 2022; Elouafi et al., 2024). It establishes a mathematical model that calculates the probability of belonging to a class, transforming the dependent variable into a binomial data type (Altun et al., 2022; Pandian et al., 2023). It is utilized for risk analysis, marketing, and social sciences (Jamil & Belkacem, 2024).

Advantages of decision trees include interpretability; they are easy to understand and visualize. The tree-like structure allows users to follow the decision-making process clearly (Almasri et al., 2022; Altun et al., 2022; Elouafi et al., 2024). They can handle both categorical and numerical variables without the need for prior transformation (Almasri et al., 2022). There is no necessity to scale or normalize data before applying a decision tree, simplifying preprocessing (Almasri et al., 2022; Elouafi et al., 2024). They are less sensitive to outliers compared to other methods like linear regression (Almasri et al., 2022; Altinay et al., 2024), can efficiently handle missing values (Elouafi et al., 2024), and require less time to train compared to neural networks (Elouafi et al., 2024; Veterini et al., 2024).

Regarding the advantages of logistic regression, it allows for rapid training even with large datasets and provides class membership probabilities, enabling clear interpretation of results and the relationship between independent variables and the dependent variable (Almasri et al., 2022; Sashank et al., 2023). It tends to be less susceptible to overfitting, especially with smaller datasets (Almasri et al., 2022). Both models are robust and can adapt to different types of data, making them useful in various predictive applications (Kocsis & Molnár, 2024; Orlando et al., 2024).

These algorithms have been applied in different areas, with education being one of the most prominent, as detailed: classification to predict student satisfaction regarding instructors (Almasri et al., 2022), classification to predict student performance in mathematics exams (Elouafi et al., 2024), prediction of academic performance and prevention of academic failures (Altun et al., 2022; Dol & Jawandhiya, 2022; Ersozlu et al., 2024; Gousia Banu et al., 2024; Tin Tin et al., 2024), prediction of student dropout (Orlando et al., 2024), evaluation of student performance and knowledge throughout their studies (Alruwais & Zakariah, 2023), analysis and evaluation of factors affecting academic performance in empirical studies (Kocsis & Molnár, 2024), prediction of students' adaptability levels in the context of online education (Grover et al., 2024), prediction of final grades (Bujang et al., 2021), and prediction of academic procrastination (Gousia Banu et al., 2024). In the health sciences, there is the prediction of depression in students (Rahman & Kohli, 2024).

However, research utilizing decision trees and logistic regression to determine student satisfaction in terms of socioeconomic level, teaching quality, relationship with instructors, course offerings, organizational climate, and campus security was not found. Therefore, this study aims to predict student satisfaction in the university educational experience using decision trees and logistic regression.

2. METHODOLOGY

To predict student satisfaction using decision trees and logistic regression, a digital questionnaire was distributed using Google Forms to all students of the National University of Altiplano in Puno, Peru. The total student population consists of 18,000 individuals. The questionnaire was sent to their institutional email addresses, resulting in a total of 1,020 responses. Since the methodology involves the application of machine learning algorithms, it was decided to work with the maximum amount of collected data to optimize the precision and robustness of the models in the analysis of student satisfaction.

The study utilized a total of 20 independent variables related to various aspects of the university educational experience. These variables encompassed factors such as socioeconomic level, quality of teaching, relationship with instructors, course offerings, organizational climate, and campus security, among others. Additionally, a dependent variable (Y) corresponding to student satisfaction was employed, with two possible outcomes: "satisfied" and "not satisfied." This structure allowed for the evaluation of the influence of the independent variables on the dependent variable Y using machine learning algorithms.

Data Collection

The questionnaire was designed to capture comprehensive information about the students' perceptions and experiences. Questions were formulated based on a thorough literature review of factors influencing student satisfaction (Ho et al., 2021; Pan & Cutumisu, 2024). The survey included demographic questions and items measured on a Likert scale to assess attitudes and satisfaction levels.

Table 1*. Independent and Dependent Variables Used in the Study*

VARIABLE CODE	VARIABLE NAME
X1	Age
X2	Sex
X3	Residence

X4	Socioeconomic Level
X5	Quality of Teaching
X6	Access to Academic Resources
X7	Relationship with Instructors
X8	Course Offerings
X9	Academic hours
X10	Quality of Facilities
X11	Access to Technology
X12	Quality of Services
X13	Organizational Climate
X14	Campus Security
X15	Extracurricular Activities
X16	Social Interaction
X17	Administrative Processes
X18	Support Services
X19	Academic Outcomes
X20	Employability
Y	Student Satisfaction

Source: Own elaboration.

Data Collection and Preprocessing

The data obtained through the Google Forms questionnaire were exported to an Excel spreadsheet for initial data cleaning. This stage involved the elimination of incomplete or inconsistent responses and the verification and correction of potential errors, ensuring the quality and reliability of the information for subsequent analysis. Data cleaning is a critical step in data analysis to enhance the accuracy of predictive models (Kotsiantis et al., 2006).

After completing the data cleaning process in Excel, the dataset was transferred to the Google Colab environment for processing. Google Colab provides a cloud-based platform with the computational capabilities necessary to efficiently execute machine learning models (Bisong, 2019). In this environment, machine learning algorithms were applied to predict student satisfaction.

Implementation of Machine Learning Algorithms

For the implementation of decision trees and logistic regression, several Python libraries were utilized, facilitating data analysis and visualization:

- Pandas and NumPy: These libraries were used for data manipulation and numerical operations, allowing efficient handling of large datasets (McKinney, 2010; Harris et al., 2020).
- Scikit-learn: This library provided the tools necessary for splitting the data into training and testing sets using the *train_test_split* function and for training the models using *DecisionTreeClassifier* and *LogisticRegression* algorithms (Pedregosa et al., 2011).
- Evaluation Metrics: The performance of the models was evaluated using *accuracy_score* to measure the proportion of correct predictions and *confusion_matrix* to visualize the performance in terms of true positives, false positives, true negatives, and false negatives.
- Visualization Libraries: Seaborn and Matplotlib were employed for visualizing the results, including generating confusion matrices and other graphical representations to interpret the model outcomes effectively (Hunter, 2007; Waskom, 2021).

Processing of Decision Trees

The processing steps for the decision tree model were as follows:

1. Data Splitting: The DataFrame was divided into two parts: predictor (independent) variables and the target (dependent) variable—student satisfaction. This step is essential to isolate the features that contribute to predicting the outcome (Kotsiantis et al., 2006).
2. Training and Testing Sets: The dataset was split into training and testing subsets, allocating 80% of the data for training the model and the remaining 20% for testing. This was done using the *train_test_split* function from Scikit-learn, ensuring that the model's performance could be evaluated on unseen data (Pedregosa et al., 2011).
3. Model Construction: A decision tree was constructed using the *DecisionTreeClassifier* algorithm. The decision tree algorithm works by recursively partitioning the data space and fitting a simple predictive model within each partition (Quinlan, 1986).
4. Model Evaluation: The model's accuracy was evaluated using the confusion matrix and the *accuracy_score* metric. The confusion matrix provided insights into the types of errors made by the model, while the accuracy score quantified the overall performance (Fawcett, 2006).

Processing of Logistic Regression

Similarly, logistic regression was implemented using the *LogisticRegression* algorithm from Scikit-learn:

1. Model Construction and Training: The logistic regression model was trained on the same training dataset, establishing a relationship between the independent variables and the probability of student satisfaction.
2. Prediction and Evaluation: Predictions were made on the test dataset, and the model's performance was evaluated using the same metrics as the decision tree model.

Ethical Considerations

Participants provided informed consent before participating in the study. All data were anonymized to protect participants' privacy, and the study complied with ethical guidelines for research involving human subjects (American Psychological Association, 2020).

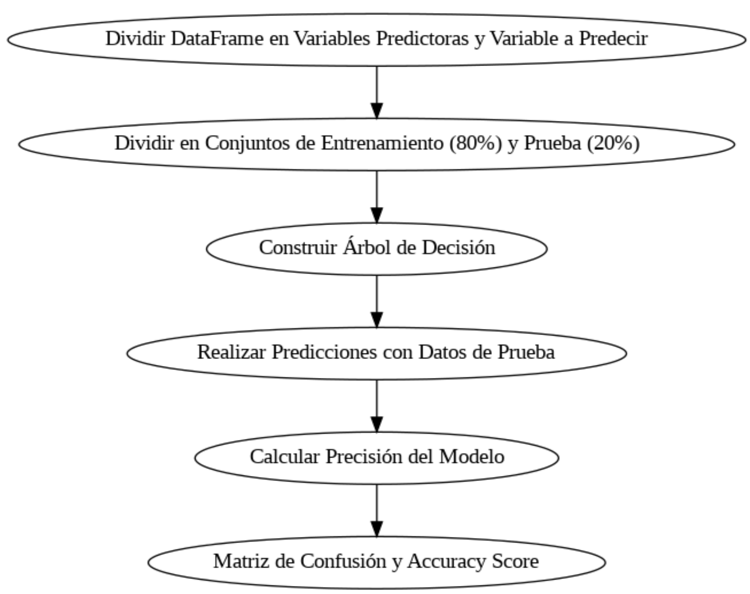

Figure 1. ***Flowchart of the Student Satisfaction Prediction Process Using Decision Trees***

Source: Own elaboration.

Logistic Regression Processing

The process began by splitting the DataFrame into two parts: the predictor variables and the target variable. The dataset was then divided into training and testing sets, with 80% of the data allocated for model training and the remaining 20% for evaluation. A logistic regression model was created, configuring the *max_iter* parameter to 1000 to ensure convergence. Subsequently, the model was trained using the training set, and predictions were made on the test set. Next, the model coefficients were obtained, which indicate the relationship between the predictor variables and the dependent variable. Finally, the model's accuracy was evaluated using the confusion matrix and the *accuracy_score* metric, providing a clear measure of the logistic regression model's performance.

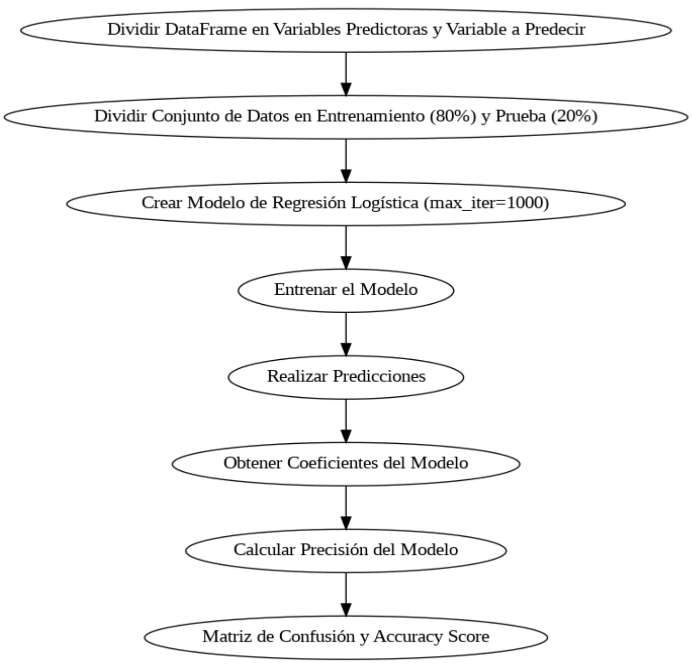

Figure 2. ***Flowchart of the Student Satisfaction Prediction Process Using Logistic Regression***

Source: Own elaboration.

Comparison of Algorithms

Finally, a comparison between the two algorithms, decision tree and logistic regression, was conducted using the confusion matrix and accuracy score as the primary evaluation metrics. These tools allowed for the analysis of each model's performance in terms of its ability to accurately predict student satisfaction. By examining the results, it was possible to determine which model offered better precision and effectiveness in predicting the dependent variable, thus facilitating an informed decision on the most suitable model for this study.

3. RESULTS

Table 2 presents the questions from the questionnaire administered to university students, which were designed to gather relevant information about their educational experience. These questions address various variables considered influential in student satisfaction, such as teaching quality, interaction with instructors, access to educational resources, and the academic environment. Collecting this data was essential for identifying key predictor variables in the decision tree and logistic regression models used in the study.

Table 2. Questionnaire Questions

VARIABLE CODE	QUESTIONS APPLIED TO EACH VARIABLE
X1	What is your age?
X2	How do you identify in terms of gender?
X3	Where do you primarily reside while studying at the university?
X4	How would you describe your household's socioeconomic level?
X5	How would you evaluate the quality of teaching you receive at the university, considering the clarity of explanations, relevance of content, and methodology used by the professors?
X6	How easy or difficult is it for you to access academic resources such as libraries, study materials, or databases necessary for your studies?
X7	How would you describe your relationship with instructors in terms of availability for consultations, academic support, and willingness to help?
X8	How diverse and relevant do you consider the course offerings to be, in relation to your interests and the demands of the professional field you are pursuing?
X9	How would you describe the academic workload of your program, considering the amount of work, course difficulty, and the time required to complete assignments and exams?
X10	How would you rate the quality of the university facilities, such as classrooms, laboratories, libraries, and common areas?
X11	How easy is it for you to access the technology required for your studies, such as computers, specialized software, and internet connection?
X12	How would you evaluate the availability and quality of services offered on campus, such as dining, transportation, medical, and recreational services?
X13	How would you describe the organizational climate at the university, including aspects such as communication, cooperation among the university community, and the general working environment?
X14	How would you evaluate the level of security on campus, considering factors like the presence of security personnel, lighting, and the overall perception of safety?
X15	How often do you participate in extracurricular activities offered by the university, such as clubs, sports, workshops, and cultural events?
X16	How would you describe the quality of your social interactions with other students at the university, considering the ease of making friends, collaborating in groups, and participating in social activities?
X17	How would you rate the efficiency of administrative processes at the university, such as enrollment, obtaining certificates, or responding to inquiries?
X18	How easy or difficult is it for you to access support services offered by the university, such as academic advising, psychological counseling, or tutoring?
X19	How satisfied are you with your academic results so far, including your grades and the learning you have achieved?
X20	What are your expectations regarding your future employability once you finish your degree, considering job opportunities in your field?
Y	Overall, are you satisfied with your educational experience at this university?

Source: Own elaboration.

To apply the decision tree model, the dataset was divided into two parts: 80% for training, corresponding to 816 data points, and 20% for testing, equivalent to 204 data points. This division allows the model to be trained on the majority of the data while using the test set to evaluate its predictive capacity and generalization on unseen data.

In the following figure, the decision tree with a maximum depth (*max_depth*) of 5 is presented, which was used to predict student satisfaction. In this tree, Class 1 represents students who are not satisfied with their educational experience, while Class 2 corresponds to students who are satisfied with their educational experience. The choice of this maximum depth seeks to balance the model's ability to capture patterns in the data without overfitting.

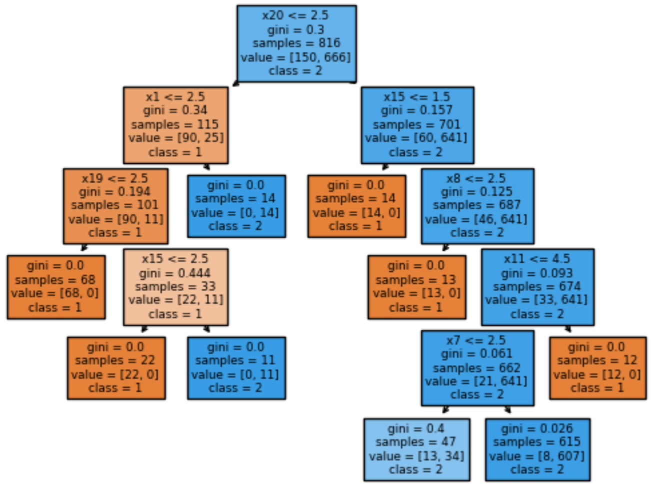

Figure 3. ***Decision Tree for Predicting Student Satisfaction***

Source: Own elaboration.

In Figure 3, it is shown that in two leaf nodes the Gini value is 0.4 and 0.026, which indicates a degree of impurity or mixture in the predictions of those nodes. In the node where the Gini index is 0.4, indicating moderate impurity, there is a mix of examples from both classes, although not equally. With a total of 47 samples (samples = 47) in this node, the value [13, 34] reflects that 13 examples belong to Class 1 (students not satisfied), and 34 belong to Class 2 (students satisfied). The dominant class is Class 2 (satisfied), as the majority of examples (34 out of 47) belong to this category, although there is still a considerable presence of examples from Class 1. This suggests that the node tends to classify students as satisfied, but with some variability.

After obtaining the decision tree graph, the model's accuracy was calculated using the confusion matrix, displayed as array([[36, 9], [0, 159]]). The confusion matrix shows that the model correctly predicted 36 students as not satisfied and made 9 errors by classifying students who were actually satisfied as not satisfied. For the satisfied students, the model was highly accurate, with 159 correct predictions and 0 errors. Therefore, the model performs excellently in predicting satisfaction (Class 2) but shows some inaccuracies in identifying students who are not satisfied (Class 1).

Using the confusion matrix, an overall accuracy of 0.956 was obtained, indicating that 95.6% of the model's predictions were correct. This reflects excellent performance in classifying student satisfaction, with only a small percentage of predictions being incorrect. This high accuracy suggests that the model is effective in identifying both satisfied and not satisfied students, although there is room for improvement, particularly in classifying students who are not satisfied.

The accuracy_score was also used to evaluate the model's performance, and an accuracy of 0.96 was achieved, indicating that 96% of the predictions made were correct. This high level of accuracy suggests that the model is very effective in classifying student satisfaction, demonstrating its ability to correctly identify both satisfied and not satisfied students.

Regarding logistic regression, similar to the decision tree, the data was split into 80% training and 20% testing. The model was created with the parameter (max_iter=1000), and the coefficients for the model were obtained.

Table 3. ***Coefficients of the Logistic Regression Model***

Variable Code	Coefficient	Variable Code	Coefficient
X1	0.082713	X11	-3.393079
X2	-0.557477	X12	1.945941
X3	-1.040448	X13	0.365841
X4	1.579024	X14	-0.078179
X5	-0.175908	X15	1.853521
X6	-0.055391	X16	-2.892646
X7	2.20463	X17	-0.885961
X8	1.244999	X18	0.170092
X9	-0.727282	X19	2.71408
X10	0.839247	X20	1.534559

Source: Own elaboration.

Student satisfaction is a crucial indicator in higher education as it is directly related to academic success, student retention, and the continuous improvement of educational institutions (Chowdhury et al., 2024; Pan & Cutumisu, 2024). Understanding the factors influencing satisfaction allows universities to make informed decisions to optimize the educational experience and, consequently, improve student performance (AL Mutawa & Sruthi, 2023). Predicting student satisfaction through decision trees is valuable due to its ability to simplify complex decisions and provide clear visual interpretations, while logistic regression is essential for quantifying the relationship between variables and accurately predicting probabilities (Anuradha et al., 2024; Hasibuan et al., 2023; Ma, 2024).

This study sought to answer the key question: What is the effectiveness of predictive models based on decision trees and logistic regression in predicting university students' satisfaction with their educational experience? The results show that both predictive models achieved high accuracy, exceeding 90%. This indicates that both decision trees and logistic regression are effective tools for identifying the most relevant factors that determine student satisfaction, enabling accurate and useful predictions for improving educational management.

The variables detailed in Table 1 were considered because they were identified as key factors in student satisfaction within the context of higher education. However, other studies have considered additional variables, such as teaching styles, the attractiveness of online learning compared to traditional learning, knowledge in the use of online learning tools (AL Mutawa & Sruthi, 2023), preference for face-to-face learning, instructors' efforts, evaluation methods (Ho et al., 2021), effectiveness, ease, and materials (Chowdhury et al., 2024), practical activities, instructor qualities, importance of preparation, and teacher behavior (Almasri et al., 2022). This difference in variable selection may be due to the particularities of the academic environment, the studied population, or the institutional approach, which could influence which aspects of the educational experience are most relevant to student satisfaction in each case.

In this study, the prediction accuracy using decision trees was 95.6% according to the confusion matrix and 96% using the accuracy_score. These values indicate robust model performance in predicting student satisfaction. Compared to other studies, such as Almasri et al. (2022), which reported an accuracy of 78.65% using decision trees, the results are superior. This difference could be due to better selection of predictor variables or higher quality data used in the analysis, reinforcing the utility of the model in this educational context. While there are not many comparable studies specifically focused on student satisfaction, the results can be contrasted with research focused on academic performance. For example, Elouafi et al. (2024) reported an accuracy of 68%, Tin Tin et al. (2024) reported 65.7%, Grover et al. (2024) reported 93.38%, Bujang et al. (2021) reported 99.1%, and Gousia Banu et al. (2024) reported 81.5% for academic performance prediction using decision trees. This comparison suggests that although both approaches aim to predict different outcomes, the accuracy of the models is superior to 65%, reflecting the robustness of this approach in educational contexts, both for academic performance and student satisfaction.

For logistic regression, the prediction accuracy was 92.6% according to the confusion matrix and 93% using the accuracy_score. Compared to other studies, such as Almasri et al. (2022), which reported an accuracy of 84.16%, and AL Mutawa & Sruthi (2023), which reported 72%, the results are superior. Although few studies focus specifically on student satisfaction, the results can be compared with research on academic performance. For example, Elouafi et al. (2024) reported an accuracy of 66%, Tin Tin et al. (2024) reported 72.9%, Grover et al. (2024) reported 77.48%, Bujang et al. (2021) reported 98.8%, and Gousia Banu et al. (2024) reported 79% for predicting academic performance using logistic regression.

In this study, the decision tree model outperformed logistic regression in both confusion matrix-based evaluation and accuracy_score. The decision tree achieved an accuracy of 95.6% according to the confusion matrix, while logistic regression obtained a slightly lower value. Similarly, in the accuracy_score, the decision tree achieved 96%, again outperforming the logistic regression model. These results highlight the greater effectiveness of decision trees in predicting student satisfaction compared to logistic regression. However, our findings do not align with those of another study conducted by Almasri et al. (2022), where logistic regression (84.16%) outperformed the decision tree (78.66%) in terms of accuracy. This discrepancy could be due to variations in the variables used, data quality, or the specific context of each investigation, suggesting that the effectiveness of each model may depend on the characteristics of the study.

On the other hand, in several studies focused on student academic performance, such as those conducted by Elouafi et al. (2024), Grover et al. (2024), Bujang et al. (2021), Rahman & Kohli (2024), Gousia Banu et al. (2024), Holgado-Apaza et al. (2023), and Selvakumar et al. (2023), decision trees demonstrated higher accuracy compared to logistic regression. These studies concluded that decision trees not only offer better predictive performance but also provide a clearer interpretation of the key variables influencing academic performance. This phenomenon relates to our study, despite being centered on student satisfaction rather than academic performance. As in the mentioned studies, where decision trees outperformed logistic regression, our research also observed higher predictive accuracy using decision trees. This consistency suggests that, regardless of the focus (student satisfaction or academic performance), decision trees may be more effective when identifying complex patterns and key variables in educational settings.

5. CONCLUSION

When comparing the effectiveness of decision tree and logistic regression models, both in terms of the confusion matrix and accuracy_score, it was found that decision trees provided better results in both cases. The decision tree's accuracy was higher, with 95.6% according to the confusion matrix and 96% in the accuracy_score, demonstrating its superior effectiveness in predicting student satisfaction compared to logistic regression, which achieved 92.6% with the confusion matrix and 93% with the accuracy_score. These findings reinforce the utility of decision trees as a more precise tool in this educational context.

The correct selection of predictor variables in this study was key to achieving high levels of accuracy in the predictive models. Both decision trees and logistic regression showed outstanding performance, with accuracies above 92%, reflecting that the selected variables effectively capture the factors influencing student satisfaction. These results underscore the importance of adequately identifying key variables to improve the predictive capacity of models and provide useful conclusions for the educational context.

The results obtained in this study offer important contributions to educational management by providing a reliable predictive model that helps identify key factors influencing student satisfaction. With an accuracy of over 95% for decision trees and 92% for logistic regression, these models can be used by universities to make data-driven decisions, improve the student experience, and optimize resources in critical areas such as teaching quality, instructor interaction, and access to resources. These findings contribute to a greater understanding of student satisfaction and allow for the development of more effective strategies to enhance student well-being and academic performance.

For future work, it is recommended to expand the study by incorporating a greater diversity of variables that may influence student satisfaction, such as emotional support and extracurricular opportunities, to further improve the predictive accuracy of the models. Additionally, it would be beneficial to conduct comparative studies in different educational contexts, such as public and private universities. Moreover, it is suggested to apply more advanced machine learning approaches, such as neural networks or random forests, which could provide higher accuracy.

REFERENCES

Abidi, S. M. R., Zhang, W., Haidery, S. A., Rizvi, S. S., Riaz, R., Ding, H., & Kwon, S. J. (2020). Educational sustainability through big data assimilation to quantify academic procrastination using ensemble classifiers. Sustainability (Switzerland), 12(15). https://doi.org/10.3390/su12156074

AL Mutawa, A., & Sruthi, S. (2023). Enhancing Human–Computer Interaction in Online Education: A Machine Learning Approach to Predicting Student Emotion and Satisfaction. International Journal of Human-Computer Interaction. https://doi.org/10.1080/10447318.2023.2291611

Almasri, A. R., Yahaya, A., & Abu-Naser, S. S. (2022). INSTRUCTOR PERFORMANCE MODELING FOR PREDICTING STUDENT SATISFACTION USING MACHINE LEARNING-PRELIMINARY RESULTS. Journal of Theoretical and Applied Information Technology, 15, 19. www.jatit.org

Alruwais, N., & Zakariah, M. (2023). Evaluating Student Knowledge Assessment Using Machine Learning Techniques. Sustainability (Switzerland), 15(7). https://doi.org/10.3390/su15076229

Altinay, F., Alpturk, K., BERİGEL, M., AYAZ, A., Dagli, G., & Altinay, Z. (2024). Evolution of Online Assessment in Higher Education: A Bibliometric Analysis. *Pakistan Journal of Life and Social Sciences*, *22*(1).

Altun, M., Kayikçi, K., & Irmak, S. (2022). A Model Proposal for Predicting Students' Academic Performances Based on Data Mining*. Hacettepe Egitim Dergisi, 37(3), 1080–1098. https://doi.org/10.16986/HUJE.2021068491

Anuradha, D., Raghavendra I, J., & Pinakapani, P. (2024). Examining Graduates' Job-Readiness Employability Skills, and Awareness using machine learning. Nanotechnology Perceptions, 20(S5), 449–454. www.nano-ntp.com

Bujang, S. D. A., Selamat, A., Ibrahim, R., Krejcar, O., Herrera-Viedma, E., Fujita, H., & Ghani, N. A. M. (2021). Multiclass Prediction Model for Student Grade Prediction Using Machine Learning. IEEE Access, 9, 95608–95621. https://doi.org/10.1109/ACCESS.2021.3093563

Chowdhury, S., Sarker, M. F. H., Khan, N. A., Islam, M. K., Tamal, M. A., & Bijoy, M. H. I. (2024). Students' Satisfaction in Online Class during COVID-19: An Empirical Study in Bangladesh Context. Education Research International, 2024. https://doi.org/10.1155/2024/1333227

Dol, S. M., & Jawandhiya, P. M. (2022). A Review of Data Mining in Education Sector. Journal of Engineering Education Transformations, 36(Special Issue 2), 13–22. https://doi.org/10.16920/jeet/2023/v36is2/23003

Elouafi, A., Tammouch, I., Eddarouich, S., & Touahni, R. (2024). Evaluating various machine learning methods for predicting students' math performance in the 2019 TIMSS. Indonesian Journal of Electrical Engineering and Computer Science, 34(1), 565–574. https://doi.org/10.11591/ijeecs.v34.i1.pp565-574

Ersozlu, Z., Taheri, S., & Koch, I. (2024). A review of machine learning methods used for educational data. Education and Information Technologies. https://doi.org/10.1007/s10639-024-12704-0

Gousia Banu, A. S., Tanveer, S., Srivani, N., Durga, M., Prasad, V., & Kumar, R. (2024). An Enhanced Machine Learning-Based Approach for Analysis and Prediction of Student Performance in Classroom Learning. In Nanotechnology Perceptions ISSN (Vol. 20, Issue S8). www.nano-ntp.com

Grover, V., Nandal, M., Balusamy, B., Sahu, D., & Dogra, M. (2024). Proposed hybrid Model in Online Education. EAI Endorsed Transactions on Internet of Things, 10. https://doi.org/10.4108/eetiot.4770

Hasibuan, M. S., Abdul Aziz, R. Z., Dewi, D. A., Kurniawan, T. B., & Syafira, N. A. (2023). Recommendation Model for Learning Material Using the Felder Silverman Learning Style Approach. HighTech and Innovation Journal, 4(4), 811–820. https://doi.org/10.28991/HIJ-2023-04-04-010

Ho, I. M. K., Cheong, K. Y., & Weldon, A. (2021). Predicting student satisfaction of emergency remote learning in higher education during COVID-19 using machine learning techniques. PLoS ONE, 16(4 April). https://doi.org/10.1371/journal.pone.0249423

Holgado-Apaza, L. A., Carpio-Vargas, E. E., Calderon-Vilca, H. D., Maquera-Ramirez, J., Ulloa-Gallardo, N. J., Acosta-Navarrete, M. S., Barrón-Adame, J. M., Quispe-Layme, M., Hidalgo-Pozzi, R., & Valles-Coral, M. (2023). Modeling Job Satisfaction of Peruvian Basic Education Teachers Using Machine Learning Techniques. Applied Sciences (Switzerland), 13(6). https://doi.org/10.3390/app13063945

Jamil, N., & Belkacem, A. N. (2024). Advancing Real-Time Remote Learning: A Novel Paradigm for Cognitive Enhancement Using EEG and Eye-Tracking Analytics. IEEE Access, 12, 93116–93132. https://doi.org/10.1109/ACCESS.2024.3422926

Kocsis, Á., & Molnár, G. (2024). Factors influencing academic performance and dropout rates in higher education. Oxford Review of Education. https://doi.org/10.1080/03054985.2024.2316616

Ma, X. (2024). Artificial intelligence-driven education evaluation and scoring: Comparative exploration of machine learning algorithms. Journal of Intelligent Systems, 33(1). https://doi.org/10.1515/jisys-2023-0319

Mazouch, P., Ptáčková, V., Fischer, J., & Hulík, V. (2018). Students who have unsuccessfully studied in the past – Analysis of causes. Journal on Efficiency and Responsibility in Education and Science, 11(3), 65–71. https://doi.org/10.7160/eriesj.2018.110303

Munshi, A., Banjar, A., Tashkandi, A., Al-Tirawi, A., Alamoudi, H., Aljojo, N., Alghoson, A., & Alsaleh, I. (2023). Arabic Sentiment Analysis for Student Evaluation using Machine Learning and the AraBERT Transformer. Engineering, Technology and Applied Science Research, 13(5), 11945–11952. https://doi.org/10.48084/etasr.6347

Orlando, J., Quispe, Q., Cuentas Toledo, O., Cuentas Toledo, M., Elio, E., Llatasi, C., Marina, E., & Saira, R. (2024). Early Prediction of University Student Dropout Using Machine Learning Models. Nanotechnology Perceptions, 20(S5), 659–669. www.nano-ntp.com

Palacios, C. A., Reyes-Suárez, J. A., Bearzotti, L. A., Leiva, V., & Marchant, C. (2021). Knowledge discovery for higher education student retention based on data mining: Machine learning algorithms and case study in chile. Entropy, 23(4). https://doi.org/10.3390/e23040485

Pan, Z., & Cutumisu, M. (2024). Using machine learning to predict UK and Japanese secondary students' life satisfaction in PISA 2018. British Journal of Educational Psychology, 94(2), 474–498. https://doi.org/10.1111/bjep.12657

Pandian, B. R., Aziz, A. A., Subramaniam, H., & Nawi, H. S. A. (2023). Exploring the role of machine learning in forecasting student performance in education: An in-depth review of literature. Multidisciplinary Reviews, 6(3). https://doi.org/10.31893/MULTIREV.2023SS043

Rahman, M. A., & Kohli, T. (2024). Mental health analysis of international students using machine learning techniques. PLoS ONE, 19(6 June). https://doi.org/10.1371/journal.pone.0304132
Raj, N. S., & Renumol, V. G. (2023). An Approach for Early Prediction of Academic Procrastination in e-Learning Environment. International Journal of Information and Education Technology, 13(1), 73–81. https://doi.org/10.18178/ijiet.2023.13.1.1782
Sashank, Y. T., Kakulapati, V., & Bhutada, S. (2023). Student Engagement Prediction in Online Session. International Journal on Recent and Innovation Trends in Computing and Communication, 11(2), 43–47. https://doi.org/10.17762/ijritcc.v11i2.6108
Selvakumar, V., Venkata, T. P., Venkata, T. P., & Singh, S. (2023). Predicting primary and middle-school students' preferences for online learning with machine learning. South African Journal of Childhood Education, 13(1). https://doi.org/10.4102/sajce.v13i1.1324
Tin Tin, T., Hock, L. S., & Ikumapayi, O. M. (2024). Educational Big Data Mining: Comparison of Multiple Machine Learning Algorithms in Predictive Modelling of Student Academic Performance Educational Big Data Mining. IJACSA) International Journal of Advanced Computer Science and Applications, 15(6). www.ijacsa.thesai.org
Veterini, A. S., Firdaus, K. M., Putri, H. S., Waloejo, C. S., Santoso, K. H., & Semedi, B. P. (2024). Photovoice as Evaluation Tool for Effective Teaching Method on the Basic Life Training. *Pakistan Journal of Life and Social Sciences*, *22*(1), 4723-4735.