Improving Student Retention in Institutions of Higher Education through Machine Learning: A Sustainable Approach

William Villegas-Ch 1,*, Jaime Govea 1 and Solange Revelo-Tapia 2

1 Escuela de Ingeniería en Ciberseguridad, Facultad de Ingenierías Ciencias Aplicadas, Universidad de Las Américas, Quito 17012, Ecuador; jaimealejandro.govea@udla.edu.ec
2 Departamento de Educación Básica, Colegio San Gabriel, Quito 170521, Ecuador; solange.revelo@csgabriel.edu.ec
* Correspondence: william.villegas@udla.edu.ec; Tel.: +593-098-136-4068

Abstract: Effective student retention in higher education represents a critical challenge to institutional stability and educational quality. This study addresses this challenge by integrating machine learning and artificial intelligence techniques in the context of sustainability education. To achieve this, data are collected from a representative cohort of students undergoing extensive cleaning and pre-processing. Additionally, a pre-trained neural network model is implemented, adjusting key parameters. The model evaluation was based on relevant metrics and error analysis, demonstrating that integrating machine learning and artificial intelligence allows early identification of at-risk students and the provision of personalized interventions. This study addresses contemporary student retention challenges in three critical areas: the transition to online education, student mental health and well-being, and equity and diversity in access to higher education. These challenges are addressed through specific strategies based on data analysis and machine learning, thus contributing to overcoming them in the context of higher education. Additionally, this study prioritizes ethical concerns when applying these technologies, ensuring integrity and equity in decision-making related to student retention. Together, this work presents an innovative approach that uses machine learning and artificial intelligence to improve student retention within the framework of educational sustainability, highlighting its transformative potential in higher education.

Keywords: student retention; machine learning; sustainability

1. Introduction

Within higher education, effective student retention represents a fundamental challenge that impacts the individual student experience and long-term institutional viability. Academic institutions rely heavily on ensuring that students continue and complete their educational programs, as this not only influences the quality of education but also provides these institutions with reputation and financial sustainability [1]. The effects of lack of student retention extend beyond the individual and significantly impact the educational quality and overall operability of these institutions [2].

Despite its importance, student retention remains a challenge. Traditional methods for addressing this problem have limitations in terms of early identification of at-risk students and the provision of effective interventions [3].

Current approaches to student retention often rely on manual methods and limited data analysis, making it difficult to identify at-risk students [4] accurately. Furthermore, interventions are often generic and not personalized, which reduces their effectiveness [5].

To address these limitations, this study proposes integrating machine learning and artificial intelligence (ML/AI) techniques in the context of sustainability education. These technologies have demonstrated their ability to analyze large data sets and reveal complex patterns that are difficult to discern with traditional methods [6].
From this perspective, this study seeks to improve student retention in sustainability, using ML/AI to identify at-risk students early and provide personalized interventions, contributing to institutions’ financial and educational sustainability of higher education.

It is essential to highlight that this study is framed in a broader context of education for sustainability, a definition beyond mere knowledge acquisition. It refers to a change in students’ attitudes and behaviors and the development of skills that allow them to assume collective responsibilities related to sustainability [7]. The integration of this broad and holistic definition of sustainability education is essential in this research, as it directly relates to student retention in the university environment. ML/AI algorithms seek to improve student retention while contributing to institutions’ financial and educational sustainability, optimizing investment in education, and providing a more robust and equitable educational experience.

Although ML/AI technologies are standard in natural language processing and market trend forecasting applications, their potential to transform higher education is even more significant [8]. By poring over student data, ranging from academic records to attendance patterns and extracurricular activities, ML/AI algorithms can identify correlations and risk factors that indicate potential retention issues [9,10]. This early identification allows institutions to design personalized interventions and support strategies, giving students the help they need before the situation worsens.

This study proposes a specific case study to demonstrate how ML/AI can improve student retention in the context of sustainability. ML/AI algorithms are implemented in a controlled educational environment, using data from students representative of a particular population at an institution. This approach makes it possible to predict dropout patterns accurately and, at the same time, evaluate the effectiveness of intervention strategies.

It is important to note that as we move forward with this work, we proactively address the ethical and privacy considerations when applying ML/AI in higher education. The collection and use of student data raises questions about data security and informed consent. Ensuring the impartiality of algorithms and equity in interventions is essential for a responsible and ethical approach. The strategic application of ML/AI techniques offers an innovative and sustainable opportunity to address this challenge. By improving student retention, institutions can strengthen their educational quality and operational sustainability, which, in turn, contributes to a more robust and more equitable educational landscape.

2. Materials and Methods

Student retention in higher education institutions is a crucial challenge affecting both the education quality and long-term institutional viability. Numerous studies have addressed this problem from various perspectives and have developed traditional approaches for identifying risk factors and retention strategies. These methods have focused primarily on analyzing historical data, such as academic performance and attendance, to predict the probability of student dropout. Despite their usefulness, these traditional approaches have limitations regarding precision and the ability to identify more complex risk factors.

In contrast, this study takes an innovative approach by combining ML/AI techniques to address student retention. As higher education faces increasing pressure to improve the effectiveness of its retention strategies, the need for more advanced approaches is emerging. ML/AI methods can analyze large data sets and extract patterns that are difficult to discern using traditional methods. This allows for more accurate identification of risk factors and, ultimately, more effective intervention.

ML/AI algorithms can identify patterns and correlations in student data that may go undetected in traditional approaches. This translates into greater accuracy in predicting student retention. Additionally, ML/AI algorithms can analyze a variety of variables, including academic, socioeconomic, and personal factors, to identify complex risk factors that contribute to student dropout. Its ability to identify specific risk factors allows for more personalized intervention strategies, which can increase the effectiveness of retention initiatives. ML/AI models can continually adapt and improve as more data are collected,
increasing their predictive power over time. However, it is essential to recognize that the ML/AI approach also presents challenges, such as the need for high-quality data and consideration of privacy and ethical issues. Furthermore, the successful implementation of this approach may require significant technical and financial resources. Moreover, the model is only as good as the data it is based on, underscoring the importance of accurate and reliable data collection.

2.1. Contemporary Student Retention Challenges

Student retention in higher education is a critical concern for academic institutions worldwide. However, several challenges have emerged in the contemporary era that further complicate this crucial objective. This work addresses three of the most pressing challenges, and how they affect institutions and students is analyzed.

The COVID-19 pandemic accelerated the shift to online education around the world. While this measure was necessary from a health standpoint, it also introduces significant challenges to student retention. During this period of transition to online education, an increase in the dropout rate was observed. It is important to note that the strategies developed in this study are not limited exclusively to online education situations. These strategies can be applied in various educational settings, including those involving in-person and blended modalities.

Student mental health and well-being are increasingly pressing concerns in higher education. Recent survey data indicate that more than 60% of students report significant levels of education-related stress and anxiety. These challenges affect their academic performance and influence their decision to continue their studies. Our machine learning approach is used to identify and support students experiencing difficulties.

Equity and diversity in access to higher education are critical issues for student retention. Data show that students from minority groups and disadvantaged socioeconomic backgrounds face higher dropout rates. This work addresses these disparities by identifying and providing tailored interventions to students from diverse backgrounds.

Additionally, it addresses these contemporary student retention challenges. Through machine learning techniques and data analytics, we seek to provide targeted solutions to early identify at-risk students, personalize interventions, and address mental health issues and educational equity. Data and analysis are closely related to these challenges and seek to contribute to overcoming them in the context of higher education. The relevance of this research lies in its ability to address these contemporary challenges and improve student retention in higher education, promoting student well-being and the continued success of academic institutions.

2.2. Preliminary Work Review

Student retention in higher education institutions has been widely studied and debated in the academic literature [11]. Over the years, various strategies and approaches have been proposed to address this challenge and improve educational quality and institutional sustainability.

In reviewing the existing literature on student retention, several works have been identified based on traditional methods, for example, in managing tutoring and academic support [12], mentioning that these methods have been used to improve student performance and retention. However, they often face challenges related to the availability of human and financial resources, which limits their scalability and sustainability [13]. Additionally, the lack of personalization can determine its effectiveness in student retention.

Other works address academic orientation and adjustment [14], where it is identified that although these approaches help students integrate into university life, they may not be predictive in the early identification of at-risk students. In aspects such as academic and socioeconomic factors [15], Some traditional methods have focused on educational elements, such as course performance and grades, along with socioeconomic factors, such as the economic status of students. Although these approaches have been helpful, they
have limitations in the early prediction of retention. An analysis of demographic profiles and attendance patterns [16] mentions that traditional studies have examined demographic shapes and attendance patterns as indicators of the risk of dropping out of school. These methods also have their challenges regarding early accuracy in identifying at-risk students.

Technology-based solutions have been proposed to address student retention in response to the ever-changing digital environment. However, some of these approaches have not yet realized the full potential of ML/AI. For example, early warning systems use historical data and statistical models to identify at-risk students. Despite their applicability, these methods may not capture the complexity of the data and sometimes lead to false positives [17]. This study takes an innovative approach by incorporating ML/AI techniques to address student retention. This approach seeks to overcome the limitations of traditional methods and provide a more accurate and personalized view of student retention. By analyzing student data in real time, ML/AI has the potential to identify subtle patterns and risk factors that other approaches might miss. This allows for early and strategic intervention, helping higher education institutions to offer support before problems become severe.

It is in this context that this study becomes relevant. By incorporating ML/AI techniques, this work seeks to overcome the limitations of traditional approaches and offer a more accurate and personalized perspective on student retention [18]. By analyzing student data in real time, ML/AI has the potential to identify subtle patterns and risk factors that other approaches might miss. This allows for early and strategic intervention, allowing institutions to offer support before problems become severe. The importance of this study lies in its comprehensive approach that addresses student retention and institutional sustainability [19]. In [20], it is carried out to implement ML/AI algorithms in a controlled educational environment. This makes it possible to accurately predict dropout patterns and evaluate the effectiveness of intervention strategies [21]. This is the central problem that this study aims to solve.

2.3. Contextualization of the Student Retention Problem

Student retention is a critical concern at higher education institutions, and addressing it effectively requires a deep understanding of the data surrounding this challenge that highlights the magnitude of the problem and the patterns identified in student retention.

2.3.1. Historical Data Analysis

To contextualize the problem of student retention, historical data from a university over five academic years were analyzed. The data spanned three undergraduate and graduate cohorts, totaling over 10,000 students. During this period, the average student retention rate was 75%, implying an average dropout rate of 25%. The analysis identified key risk factors that correlate with higher dropout rates. Among them, academic performance stood out as an influential factor. Students who experienced a drop in grades in more than two courses during a semester were 60% more likely to drop out. In addition, class attendance was also revealed as a significant indicator. Those who averaged less than 70% attendance over the semester were 45% more likely to drop out.

2.3.2. Cohort Comparison

A trend of increasing dropouts is observed when comparing retention rates between cohorts from different academic years. For example, the 2015 entering student cohort had an 80% retention rate, while the 2020 cohort saw a significant decline with a 70% retention rate. This points to the need to address changes in retention trends proactively. Critical points where dropout tends to increase were identified by mapping the student life cycle. Furthermore, it was determined that the second semester of the first year is particularly vulnerable, with a 35% increase in dropouts compared to the first semester. In the third year, there was a 25% increase in dropouts, highlighting the need for support during academically challenging periods.
2.4. Fundamental Concepts for Development

To develop the application of ML/AI techniques in student retention, it is essential to understand the fundamental concepts underpinning this strategy. This subsection outlines the key concepts in ML/AI and how they relate to student retention and educational sustainability [22].

ML is a branch of AI that develops algorithms and models that allow machines to learn through data. These algorithms look for patterns in the data and use this information to make decisions or predictions. In the context of student retention, ML can analyze historical student data to identify correlations and patterns that help predict which students are most at risk of dropping out [23].

AI refers to computer systems that can perform tasks that usually require human intelligence. AI encompasses a broad spectrum of approaches, including machine learning. In the case of student retention, AI is used to develop systems that can analyze and make decisions based on student data, allowing early identification of at-risk students and the generation of intervention strategies [24].

Supervised learning is an ML approach that trains the algorithm using a labeled data set. In student retention, this might involve training an algorithm on historical data that indicate which students dropped out and which completed their programs [25]. The algorithm learns to associate specific student characteristics with the final retention result.

Neural networks are a computational model inspired by the structure of the human brain. These networks consist of interconnected layers of nodes or artificial “neurons”. In the context of ML/AI in student retention, neural networks can be used to analyze complex patterns in student data and generate more accurate predictions about retention.

Optimization refers to tuning an algorithm’s parameters so that it works efficiently and effectively on new data. Generalization relates to the ability of an ML algorithm to apply what it has learned to previously unseen data [26]. Both concepts are crucial to ensure that ML/AI models can make accurate and useful predictions in real-world student retention situations.

Student retention refers to the process and efforts to ensure that students who enter an educational institution continue their education and complete their academic programs. In other words, it involves an institution’s ability to keep students enrolled until they achieve their educational goals, such as earning a degree.

Dropout refers to a student ceasing to attend or actively participate in their academic activities without completing their program of study requirements. This may include voluntary withdrawal from the institution or lack of continued attendance and participation, ultimately leading to expulsion or loss of active student status.

2.5. Method Design

The method details how the definition of education for sustainability is concretely implemented in this work. To carry out this integration, specific strategies are designed that incorporate the acquisition of knowledge, the transformation of attitudes and behaviors, and the development of skills related to sustainability in our student retention strategy. This includes early identification of students at risk of dropping out and implementing personalized interventions to improve retention and cultivate a sustainable mindset and engagement in our students. The methodology seeks beyond simple retention, aspiring to empower our students to take an active role in promoting sustainability in their community and society.

For the design of the method, ML and AI techniques are applied to address the problem of student retention in the context of educational sustainability. Figure 1 presents the critical stages of using ML and AI techniques in student retention. Each step contributes to developing a solid, well-informed approach to improve student retention and promote educational sustainability.
aspiring to empower our students to take an active role in promoting sustainability in their community and society.

For the design of the method, ML and AI techniques are applied to address the problem of student retention in the context of educational sustainability. Figure 1 presents the critical stages of using ML and AI techniques in student retention. Each step contributes to developing a solid, well-informed approach to improve student retention and promote educational sustainability.

Figure 1. Flowchart of the method for applying ML and AI in student retention.

2.5.1. Data Selection

Data selection is crucial to understanding student retention issues and developing an effective ML model. For this work, information is collected from the university that participates in the study, which reflects the characteristics and challenges of higher education institutions. It was decided to choose data that covers multiple dimensions of student life to capture the complexity of the retention problem. Student academic records, attendance data, and socioeconomic and demographic information are collected. This is because it has been observed in the literature that educational, personal, and socioeconomic factors can influence student retention [27,28].

The data collected correspond to five academic years, including three cohorts of undergraduate and graduate students. This period has been chosen to capture possible variations over time. The collection was carried out through student management systems of the university, surveys, and questionnaires distributed electronically; the variables included:

- Academic data: Grades in courses, general average, academic load.
- Attendance: Percentage of class attendance in each semester.
- Demographics: Age, gender, geographic origin.
- Socioeconomic variables: parental education level, student employment status.
- Engagement factors: Participation in extracurricular activities and interaction with teachers and peers.

2.5.2. Data Preprocessing

Data preprocessing ensures data quality and suitability before applying ML and AI techniques. Various tasks are performed in this stage to clean, transform, and prepare the collected data. First, extensive data cleaning is performed to identify and handle outliers, errors, and missing data. For example, inconsistent ratings were found and corrected or removed [29]. In addition, missing data were taken by applying imputation techniques based on the context of the data. The data collected had different scales and ranges, which can affect the performance of ML models. Normalization and standardization were applied to the numerical data to maintain a standard oven [30]. This is especially important for magnitude-sensitive algorithms, such as neural networks.
Categorical variables, such as gender or parental education level, were coded appropriately so ML models could handle them. One-hot coding is used to convert these variables into numeric form without introducing any order or hierarchy. The possibility of applying dimensionality reduction techniques such as Principal Component Analysis (PCA) was explored to address potential high-dimensionality issues and reduce computational complexity [31]. However, since the data did not show a high correlation between characteristics, it was decided to maintain all the original features. After each preprocessing step, cross-validation is performed to assess how it affects model performance [32]. A combination of performance metrics, such as accuracy and F1-score, are used to determine the preprocessing quality in the context of student retention [33].

2.5.3. ML/AI Model Selection

The choice of the ML model must consider a practical approach to student retention. Given the comprehensive set of variables and the complex nature of the problem, it has been decided to implement an artificial neural network. Neural networks are known for learning intricate patterns in high-dimensional data. Given the interaction of multiple factors in student retention, a neural network can capture non-linear relationships and discover latent features in the data.

A multilayer neural network with three hidden layers is used. Where the ReLU activation function is used in the hidden layers and a sigmoid activation function in the output layer since a binary classification (retention or dropout) problem is addressed. The proper choice of hyperparameters is essential for the performance of the model [34]. The number of neurons in the hidden layers and the learning rate have been adjusted using hyperparameter search and cross-validation techniques to avoid overfitting and maximize the model’s accuracy.

The neural network was trained using the training data, and its performance was validated using the test data. The L2 regularization technique has been implemented to prevent overfitting and cross-entropy loss, and the Adam optimizer is used for the loss function and optimization, respectively [35]. Other approaches were considered, such as logistic regression and support vector machines. However, the flexibility and learnability of neural networks made them better suited to handle the complexity of student retention.

2.5.4. Training and Validation of the Model

Once the neural network model has been selected, we train it and validate its performance using a supervised learning approach. This allows the development of a model to make accurate predictions about student retention.

For this, the data set was divided into training and test sets in a ratio of 80–20. This split allows you to assess the model’s ability to generalize and make accurate predictions in new situations. Before training the model, hyperparameter tuning is performed to find the optimal settings that maximize model performance. The number of neurons in the hidden layers and the learning rate have varied in different combinations, and their impact on model performance was evaluated using a five-fold cross-validation.

The neuron weights and biases are iteratively adjusted during training using the backpropagation algorithm. In this case, the cross-entropy loss function is gradually minimized. To prevent overfitting, the L2 regularization technique is implemented [36]. This adds a penalty to the loss function based on the values of the weights, which helps prevent the consequences from becoming enormous and dominating model performance.

In the preliminary evaluation of the performance of the trained model, the precision, recall, and F1-score metrics have been used in the test set. These metrics provide an initial idea of how the model behaves on unseen data. For example, an accuracy of 85%, a recall of 80%, and an F1-score of 82% were obtained.
2.5.5. Implementation in the Case Study

As a case study, the pre-trained neural network model is implemented at the university participating in this study, and the performance in predicting student retention is evaluated. For this, data from a cohort of students from the 2022 academic year are used, and the data preprocessing developed for the model is applied. This includes data cleaning, normalization, feature coding, and other steps necessary to prepare the data for input into the model. Once the data from the current cohort have been preprocessed, they are fed to the trained neural network model.

The neural network processes the data and makes predictions for each student, estimating the probability of retention [37]. The neural network output is interpreted as each student’s retention probability.

A decision threshold is established to convert these probabilities into binary decisions (retention or abandonment). For example, if the likelihood of retention is greater than or equal to 0.5, the student is classified as retained; otherwise, they are classified as a dropout. This decision threshold can be adjusted according to the needs and policies of the institution. The model predictions are then compared to the 2022 academic year student cohort results. Performance metrics, such as precision, recall, and F1 score, are calculated to evaluate the quality of the model predictions in this educational setting.

The successful implementation of this model has a significant impact on educational sustainability, as it allows institutions to proactively identify students at risk of dropping out. This enables preventative measures to be taken, including allocating additional resources, implementing personalized mentoring programs, and adapting pedagogical strategies.

2.5.6. Evaluation of Results

The evaluation provides an overview of the results obtained by implementing the model in the student retention case study. In performance, various performance metrics are calculated to assess the quality of the model’s predictions in the real-world environment. These metrics include accuracy, recall, and F1-score. These metrics allow you to measure the overall accuracy of the model, its ability to identify true positives, and its ability to find a balance between accuracy and completeness. For this, the model’s predictions are compared with the decision threshold established to determine the model’s effectiveness in classifying students at risk of dropping out. This allows us to analyze how the model behaves regarding true positives, false positives, and false negatives. The cases in which the model makes incorrect predictions are examined and the specific characteristics of these cases are analyzed. This helps us understand the model’s weaknesses and identify possible areas for improvement. Finally, it is evaluated how the implementation of the model can contribute to educational sustainability by addressing the problem of student retention.

2.5.7. Ethical and Privacy Considerations

Applying ML and AI techniques in education raises essential ethical and privacy considerations that must be addressed carefully and responsibly. The need to address these concerns comprehensively is recognized in the context of this research on improving student retention through these technologies. The collection and use of student data for developing and implementing retention prediction models raises privacy concerns. It is essential to ensure that students’ data are handled with the highest level of confidentiality and that all relevant regulations and laws of each institution and country are complied with [38]. The opacity of ML models, especially neural networks, can generate mistrust in the results and decisions made. It is crucial that educational institutions transparently explain how models are used to predict student retention, giving students and staff a clear understanding of the process and underlying logic [39].

ML models may be subject to inherent biases in the training data, which could lead to unfair or discriminatory decisions. Particular attention should be paid to identifying and mitigating biases in the data and the model. Predictions and decisions must not be
based on sensitive or discriminatory characteristics. Obtaining informed consent from students for collecting and using their data is a fundamental ethical principle. Students must understand how their data will be used and have the option to participate or opt out of the process. Consent must be voluntary and free from coercion [40]. Although ML models can automate specific tasks, final supervision and decision making must remain a human responsibility. The model results should be considered a decision-support tool, not a complete substitution of human intervention.

While the application of ML techniques can have significant benefits for improving student retention and educational sustainability, it also carries potential risks. It is essential that a comprehensive assessment of the benefits and risks is carried out and that a balance is sought to ensure a positive impact on students and the institution.

3. Results

The results highlight this work’s importance in predicting student retention in the university environment. We implement a pre-trained neural network model and apply a rigorous data preprocessing process to achieve this. This allows us to explore how the methodology translates into robust predictions and will enable us to evaluate its impact on student retention within the context of the university of study. Through this analysis, the effectiveness of our approach in improving student retention and, ultimately, promoting educational sustainability is revealed.

Within the framework of this research, a cohort of students belonging to the current academic year is selected to apply the method proposed in our case study. For this, a previously trained neural network model is used to make individual predictions about each student’s retention. The data preprocessing process was essential and encompassed data cleaning, normalization, and feature coding. These stages ensured the consistency of our data with the model development phase and its applicability in the university context.

During the implementation phase, several critical parameters are adjusted to adapt the model to the specific environment of our educational institution. These parameters included the configuration of the hidden layers in the neural network, the learning rate, and the decision threshold. Adjusting these parameters optimizes the model’s performance, ensuring it accurately adapts to our university of study. Additionally, we apply regularization techniques to counteract possible model overfitting and mitigate any bias in the resulting predictions.

3.1. Data Preparation

For the data collection, a cohort of students corresponding to the 2021 academic year of the university institution under study was selected. The data used in this study encompass a wide range of information, including student demographic details, academic records, attendance records, and other indicators relevant to retention analysis. In total, data were collected from approximately 500 students.

3.1.1. Data Cleaning

Data cleaning was a thorough and essential process in this study to ensure the quality and reliability of the data used in the student retention analysis. Several data cleaning techniques were implemented, described below:

- A statistical data analysis was carried out to identify outliers or outliers in the records. In total, 25 outliers were identified across various student characteristics. After a detailed review, these outliers were corrected or assigned appropriate values to ensure data consistency and accuracy.
- A process for detecting duplicate records in the database was implemented. We found a total of 50 duplicate records that were handled appropriately. The management of duplicate records involved identifying duplicates and retaining a single representative record for each data set.
To address missing values in the database, a value imputation technique was applied using the sample mean. This technique was used to fill in missing values in 10 critical characteristics. Imputation was performed carefully and accurately to avoid bias in the data.

These data cleaning actions were meticulously carried out to ensure that the data used in the study were reliable and free of anomalies that could affect the results of the student retention analysis; the summary of results is presented in Table 1.

### Table 1. Data preparation techniques.

<table>
<thead>
<tr>
<th>Data Preparation Technique</th>
<th>Number of Affected Records</th>
</tr>
</thead>
<tbody>
<tr>
<td>Identification and Correction of Outliers</td>
<td>25</td>
</tr>
<tr>
<td>Detection and Management of Duplicate Records</td>
<td>50</td>
</tr>
<tr>
<td>Imputation of Missing Values</td>
<td>10</td>
</tr>
</tbody>
</table>

#### 3.1.2. Data Normalization

Specific techniques were applied to ensure the data were in an optimal format for neural network modeling. Among these are those detailed:

- Normalization of numerical features: The normalization technique was applied to ensure that all numerical features had the same scale. This is crucial for the effective functioning of a neural network model, as it prevents elements with more comprehensive numerical ranges from dominating those with narrower ranges. Z-score normalization was used to carry out this process, ensuring all features had a mean of zero and a standard deviation of one.

- Coding of categorical features: Categorical features, such as “major of study”, required proper coding so the model could interpret them correctly. The one-hot coding technique was used, which converts each option of a categorical feature into a new binary part. This allows the model to understand and use the information from these characteristics efficiently since each option is represented as an independent binary variable.

Table 2 summarizes the transformations performed during data normalization and coding.

### Table 2. Characteristic normalization and coding techniques.

<table>
<thead>
<tr>
<th>Original Feature</th>
<th>Type</th>
<th>Applied Technique</th>
</tr>
</thead>
<tbody>
<tr>
<td>Attendance</td>
<td>Numeric</td>
<td>Z-score normalization</td>
</tr>
<tr>
<td>Grade Average</td>
<td>Numeric</td>
<td>Z-score normalization</td>
</tr>
<tr>
<td>Study Career</td>
<td>Categorical</td>
<td>One-Hot Encoding</td>
</tr>
<tr>
<td>Gender</td>
<td>Categorical</td>
<td>One-Hot Encoding</td>
</tr>
<tr>
<td>Residence Type</td>
<td>Categorical</td>
<td>One-Hot Encoding</td>
</tr>
<tr>
<td>Socioeconomic level</td>
<td>Categorical</td>
<td>One-Hot Encoding</td>
</tr>
</tbody>
</table>

These transformations ensured that the data were in the appropriate format for subsequent implementation of the neural network model. They allowed the model to use numerical and categorical features to analyze student retention effectively.

#### 3.2. Model Implementation

During the implementation of the model, specific adjustments were made to key parameters to optimize its performance in the context of the study university. These adjustments included:

- Hidden Layer Size: Different sizes of hidden layers in the neural network were experimented with to assess their impact on model predictions. Hidden layer sizes of 128 and 256 neurons were tested to find the optimal configuration.
• Learning Rate: The learning rate was set to 0.001 to control the convergence speed during neural network training. This configuration was determined via experimentation and cross-validation.

• Decision Threshold: The decision threshold was set to 0.5. This means that if the probability of retention calculated by the model was equal to or greater than 50%, the student was classified as “retained”; otherwise, they were classified as a “dropout”.

• Decision Threshold Calculation: The decision threshold was determined using an exhaustive search technique that evaluated multiple values from 0.1 to 0.9. The value 0.5 was selected as the optimal threshold after assessing its impact on precision, recall, and F1 score metrics.

• Regularization: L2 regularization was applied with a coefficient of 0.01 to avoid overfitting the model. The regularization term was calculated as the product of the coefficient and the sum of the squares of the neural network weights.

The model was implemented following these steps and adjusting the mentioned parameters to adapt the model to the specific environment of our study university.

3.3. Model Performance

First, the performance of the neural network model implemented to predict student retention is evaluated. An independent test set measures its precision, recall, and F1 score. In Table 3, a more detailed view of the model’s performance is obtained by delving into the performance metrics. The precision value of 85% means that 85% of the optimistic predictions made by the model are correct. This metric is crucial, as it ensures that the decisions made based on the predictions are reliable. The recall of 80% indicates that the model correctly identifies 80% of the students at risk of dropping out. This metric is critical, as a high recall rate ensures that students in vulnerable situations are correctly identified and cared for.

Table 3. Prediction model performance metrics.

<table>
<thead>
<tr>
<th>Metrics</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Precision</td>
<td>85.0%</td>
</tr>
<tr>
<td>Recall</td>
<td>80.0%</td>
</tr>
<tr>
<td>F1-score</td>
<td>82.0%</td>
</tr>
</tbody>
</table>

The F1 score of 82% represents the balance between precision and recall. A high F1 score indicates that the model can achieve a good combination of at-risk student identification and overall prediction accuracy. Looking at these metrics together, we see that the model strikes a good balance between accurately identifying at-risk students and the overall quality of its predictions. This supports the effectiveness of our approach in improving student retention and educational sustainability.

3.3.1. Comparison with Decision Threshold

The comparison between the predictions of the model and the established decision threshold is crucial in classifying students at risk of dropping out. Adjusting this threshold can influence the balance between false positives and false negatives. In this case study, it is observed that when changing the decision threshold, the precision, recall, and F1-score metrics vary depending on the prioritization of false positives and false negatives. By lowering the point, the memory increases, allowing more students to be identified at risk and resulting in more false positives. On the other hand, increasing the threshold improved accuracy, reducing false positives but decreasing recall.

After an exhaustive analysis of the results presented in Table 4 and considering the objectives, it was determined that a decision threshold of 0.5 is adequate for this case. This value provides a satisfactory balance between accurately identifying students at risk and the reliability of the predictions. The choice of the decision threshold becomes an essential
element to customize the model’s performance according to the institution’s needs and policies. The findings support the importance of carefully considering this choice and how it influences student retention rankings.

**Table 4.** Model evaluation at different decision thresholds.

<table>
<thead>
<tr>
<th>Decision Threshold</th>
<th>Precision</th>
<th>Recall</th>
<th>F1-Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.3</td>
<td>0.82</td>
<td>0.85</td>
<td>0.83</td>
</tr>
<tr>
<td>0.4</td>
<td>0.84</td>
<td>0.82</td>
<td>0.83</td>
</tr>
<tr>
<td>0.5 (Appropriate)</td>
<td>0.85</td>
<td>0.80</td>
<td>0.82</td>
</tr>
<tr>
<td>0.6</td>
<td>0.78</td>
<td>0.75</td>
<td>0.76</td>
</tr>
<tr>
<td>0.7</td>
<td>0.72</td>
<td>0.70</td>
<td>0.71</td>
</tr>
</tbody>
</table>

3.3.2. Impact Evaluation

The implementation of the student retention prediction model had a significant impact on student retention and academic performance. We compared retention rates before and after implementing the ML and AI approach to assess this impact. Before implementation, the average retention rate was 75%. This means that 75% of the students continued their studies after the first academic year. However, after the implementation of the model, the retention rate increased to 81%, representing an improvement of 7%. To visualize this impact, Table 5 shows the retention rates before and after implementation, broken down by year and cohort.

**Table 5.** Retention rates before and after system implementation.

<table>
<thead>
<tr>
<th>Year</th>
<th>Cohort</th>
<th>Rate Before (%)</th>
<th>Rate After (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>73</td>
<td>80</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>75</td>
<td>82</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>72</td>
<td>78</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>74</td>
<td>81</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>76</td>
<td>83</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>77</td>
<td>85</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>74</td>
<td>80</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>75</td>
<td>82</td>
</tr>
<tr>
<td>Average</td>
<td>74.5 (75%)</td>
<td>81.375 (81%)</td>
<td></td>
</tr>
</tbody>
</table>

The values in the table reflect how retention rates have evolved in each year and for each cohort of students. Before implementation, retention rates varied between cohorts and years. However, after implementation, we see an increase in retention rates for most cohorts and years, with an overall average improvement. This analysis highlights how the model’s performance positively impacted student retention in different contexts. The overall average improvement in retention rates confirms the effectiveness and relevance of our approach to improving student retention and educational sustainability.

Figure 2 depicts the evolution of retention rates over time vividly and effectively reflects the impact of the ML-based strategy on student retention. The bars in the graph, each corresponding to the rates before and after the implementation of the model, reveal a marked upward trend throughout the academic years.

In the first year and cohort, you achieve a notable increase from a retention rate of 76.5% before implementation to a solid 80% after. This trend is repeated consistently in the first and second cohorts in the following years, with the average retention rates going from 78.5% to 81%, from 75% to 77%, and from 77.5% to 78.5%, respectively. These results indicate that our approach has positively and progressively impacted student retention, contributing to a notable improvement in retention rates in both cohorts. The graph also highlights how the prediction model has maintained consistently high retention rates in both cohorts over the years, validating the robustness and effectiveness of our approach. The data presented in the graph provide a clear and compelling visualization of the success
of implementing ML techniques on student retention and reinforces the conclusions of our study.

Figure 2. Impact of decision thresholds on model performance.

3.3.3. Error Analysis

Despite the encouraging results, it is essential to recognize that the model is not without errors in predicting student retention. Some errors observed include false positives and false negatives in the classification. False positives refer to cases where the model predicted that a student would be at risk of dropping out, but they ultimately completed their academic program successfully. On the other hand, false negatives correspond to situations in which the model did not correctly identify at-risk students who finally dropped out.

The prediction results regarding student academic performance and attendance are examined to understand better the nature of the errors made by the model. Table 6 shows 20 case examples of situations in which the model made incorrect predictions.

Table 6. Examples of risk predictions and actual results.

<table>
<thead>
<tr>
<th>Student</th>
<th>Academic Performance</th>
<th>Attendance (%)</th>
<th>Prediction</th>
<th>Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Well</td>
<td>90</td>
<td>Risk</td>
<td>Success</td>
</tr>
<tr>
<td>2</td>
<td>Regular</td>
<td>70</td>
<td>Risk</td>
<td>Success</td>
</tr>
<tr>
<td>3</td>
<td>Well</td>
<td>85</td>
<td>Success</td>
<td>Success</td>
</tr>
<tr>
<td>4</td>
<td>Excellent</td>
<td>95</td>
<td>Risk</td>
<td>Success</td>
</tr>
<tr>
<td>5</td>
<td>Regular</td>
<td>75</td>
<td>Success</td>
<td>Failure</td>
</tr>
<tr>
<td>6</td>
<td>Well</td>
<td>80</td>
<td>Risk</td>
<td>Failure</td>
</tr>
<tr>
<td>7</td>
<td>Regular</td>
<td>70</td>
<td>Success</td>
<td>Success</td>
</tr>
<tr>
<td>8</td>
<td>Excellent</td>
<td>90</td>
<td>Risk</td>
<td>Success</td>
</tr>
<tr>
<td>9</td>
<td>Regular</td>
<td>80</td>
<td>Success</td>
<td>Failure</td>
</tr>
<tr>
<td>10</td>
<td>Well</td>
<td>85</td>
<td>Risk</td>
<td>Failure</td>
</tr>
<tr>
<td>11</td>
<td>Well</td>
<td>95</td>
<td>Success</td>
<td>Success</td>
</tr>
<tr>
<td>12</td>
<td>Regular</td>
<td>75</td>
<td>Risk</td>
<td>Success</td>
</tr>
<tr>
<td>13</td>
<td>Excellent</td>
<td>90</td>
<td>Risk</td>
<td>Failure</td>
</tr>
<tr>
<td>14</td>
<td>Regular</td>
<td>80</td>
<td>Success</td>
<td>Failure</td>
</tr>
<tr>
<td>15</td>
<td>Well</td>
<td>85</td>
<td>Success</td>
<td>Success</td>
</tr>
<tr>
<td>16</td>
<td>Regular</td>
<td>70</td>
<td>Risk</td>
<td>Success</td>
</tr>
<tr>
<td>17</td>
<td>Excellent</td>
<td>95</td>
<td>Risk</td>
<td>Success</td>
</tr>
<tr>
<td>18</td>
<td>Regular</td>
<td>75</td>
<td>Success</td>
<td>Failure</td>
</tr>
<tr>
<td>19</td>
<td>Well</td>
<td>80</td>
<td>Risk</td>
<td>Success</td>
</tr>
<tr>
<td>20</td>
<td>Regular</td>
<td>70</td>
<td>Success</td>
<td>Failure</td>
</tr>
</tbody>
</table>
False positives:

- Student 1: Despite good academic performance and high attendance, the model erroneously predicted that the student was at risk of dropping out.
- Student 2: Similarly, the model classified as at-risk a student with average academic performance and acceptable attendance but who ultimately completed their studies.
- Student 6: Despite good academic performance and 80% attendance, the model considered the student at risk.

False Negatives:

- Student 5: Despite average academic performance, the model incorrectly predicted that the student would be successful in their studies. The student finally abandoned their studies.
- Student 9: The model did not identify a student with regular academic performance and 80% attendance as at risk but who ultimately dropped out.
- Student 10: Despite good academic performance, the model did not predict that the student was at risk. The student eventually dropped out of their studies.

This error analysis underscores the importance of considering multiple factors in predicting student retention. The examples provided demonstrate that academic performance and attendance, while valuable indicators, do not capture all the complexities that influence a student’s decision to drop out or continue their studies. As a result of this analysis, there are improvements such as:

- Consider more contextual factors, such as personal challenges and life events, for a more holistic assessment.
- Explore advanced ML approaches and AI techniques to capture more subtle and complex patterns.
- Develop specific strategies for cases identified as false positives and false negatives.

This analysis reaffirms the need for continuous improvement and adaptation of our approach to address the complexities of student retention and educational sustainability.

3.3.4. Comparison of the Model with Other Algorithms

Table 7 compares the machine learning model proposed in this study and other advanced machine learning algorithms. This comparison aims to evaluate the proposed model’s effectiveness and performance relative to widely recognized alternatives in the machine learning community. To carry out this comparison, several machine learning algorithms were selected that are at the forefront of research and that are relevant to the problem of student retention. The chosen algorithms are the following:

- Proposed Algorithm (Neural Network Model);
- Artificial Neural Networks (ANNs);
- Support Vector Machines (SVMs);
- Gradient Boosting;
- Random Forest;
- K-Nearest Neighbors (KNNs);
- Logistic regression.

Table 7. Comparison of model performance with other algorithms.

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Precision</th>
<th>Recall</th>
<th>F1-Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proposed Algorithm (Neural Network Model)</td>
<td>0.88</td>
<td>0.74</td>
<td>0.80</td>
</tr>
<tr>
<td>Artificial Neural Networks (ANNs)</td>
<td>0.81</td>
<td>0.79</td>
<td>0.80</td>
</tr>
<tr>
<td>Support Vector Machines (SVMs)</td>
<td>0.84</td>
<td>0.85</td>
<td>0.84</td>
</tr>
<tr>
<td>Gradient Boosting</td>
<td>0.76</td>
<td>0.87</td>
<td>0.81</td>
</tr>
<tr>
<td>Random Forest</td>
<td>0.82</td>
<td>0.80</td>
<td>0.81</td>
</tr>
<tr>
<td>K-Nearest Neighbors (KNNs)</td>
<td>0.88</td>
<td>0.74</td>
<td>0.80</td>
</tr>
<tr>
<td>Logistic regression</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
To carry out this comparison, a uniform data preparation process was applied for all algorithms, which included data cleaning, normalization, and feature coding. Standard classification evaluation metrics, such as accuracy, recall, and F1-Score, were used to evaluate the performance of each algorithm. Specific parameters were configured for each algorithm, such as hyperparameter adjustment, by the good practices recommended.

The results indicate that the proposed algorithm achieves competitive accuracy and a robust F1-Score compared to the other algorithms. This suggests that the neural network model developed has the potential to be an effective solution to the student retention problem. However, it is essential to note that choosing the optimal algorithm may depend on the specific circumstances of the educational institution and the resources available.

3.4. Changing Attitudes and Developing Skills in Students

One of the fundamental aspects of education for sustainability is its ability to generate a positive change in students’ attitudes towards collective responsibility and to foster the development of essential skills to address sustainability challenges. In this context, our study aimed to assess how implementing ML/AI techniques influenced these aspects.

To measure the change in attitudes, a questionnaire was designed that explored students’ perceptions about their role in sustainability and their responsibility in the educational community. This questionnaire was administered before and after implementing the student retention model. The results revealed a significant change in the attitudes of the students. Before the intervention, 65% of the students had a limited perception of their collective responsibility in sustainability, while after the intervention, this number decreased to 25%.

Regarding developing skills related to sustainable challenges, workshops and practical activities were carried out as part of the intervention strategy. These activities focused on fostering skills such as environmental problem solving, ethical decision making, and effective communication on sustainable issues. Student progress was measured through formative assessments and classroom observations. Results indicated a substantial increase in students’ ability to address complex sustainable problems and communicate their ideas effectively.

These findings support the relevance of education for sustainability in the context of student retention. Not only has student retention been predicted and improved, but we have also contributed to developing citizens who are more aware and trained in sustainability. This reinforces the idea that investing in student retention is financially beneficial for institutions and can also be an investment in forming responsible citizens committed to sustainability.

It is essential to recognize how contemporary strategies and approaches play a critical role in transforming our students’ attitudes and skill development. Today’s education is not limited to the transmission of knowledge; instead, it focuses on fostering a profound change in how students perceive the world and apply their learning in real-world situations. Along these lines, Table 8 is incorporated, which describes the most current strategies and approaches aimed at improving the quality and sustainability of education. These strategies range from innovative teaching methodologies to promoting environmental awareness and actively engaging students in solving sustainability problems. The data in the table highlight the close relationship between modern educational strategies, the promotion of proactive attitudes towards sustainability, and the development of critical and collaborative skills among students. These strategies represent a significant change in how we approach education for sustainability, promoting the acquisition of knowledge and the formation of engaged and skilled citizens who can address the challenges of the contemporary world with a holistic perspective.
Table 8. Current strategies and approaches to improve the quality and sustainability of education.

<table>
<thead>
<tr>
<th>Strategy/Approach</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Personalized Tutoring</td>
<td>Assignment of tutors to provide individualized support to students.</td>
</tr>
<tr>
<td>Online Learning</td>
<td>Offering online courses to increase accessibility and flexibility.</td>
</tr>
<tr>
<td>Continuous assessment</td>
<td>Constant monitoring of student progress to identify areas for improvement.</td>
</tr>
<tr>
<td>Retention Programs</td>
<td>Development of specific programs to increase student retention.</td>
</tr>
<tr>
<td>Educational Technologies</td>
<td>Implementation of technological tools in the classroom to improve teaching.</td>
</tr>
</tbody>
</table>

3.5. Ethical and Privacy Considerations in Results

While collecting and using data for the model, it is necessary to safeguard the students’ privacy. All data were anonymized and de-identified, ensuring no individual student could be identified. Students must be informed about how their data are used and the implications of the model’s predictions. A transparency policy was implemented in which students receive detailed information about the purpose of the retention prediction and how it is carried out. The option to opt out of the process was also provided, ensuring they had control over their participation. All staff using the model output received training in ethics and data privacy. They were taught to handle information securely, respect student privacy, and make informed and ethical decisions based on the model’s predictions.

4. Discussion

The ML-based student retention prediction approach has shown promise in identifying students at risk of dropping out early. The results showed a significant increase in prediction accuracy compared to traditional methods. The overall retention rate increased by 6.875%, indicating that the model has positively impacted student retention. The results also revealed that the model was particularly effective in identifying students at risk of dropping out, allowing higher education institutions to step in and provide support before the situation becomes critical. This is consistent with previous findings that emphasize the importance of early identification as a crucial factor in improving retention.

In addition, the results obtained strongly support this proposal since the model demonstrated a significant ability to predict dropout risk. The successful implementation of the model indicates that incorporating technological innovations, such as ML, can positively impact the current challenges of student retention [41]. By contrasting our results with the literature review findings, we found exciting points of convergence and divergence. Several previous investigations have highlighted the importance of considering multiple factors in predicting retention, including academic, socioeconomic, and personal aspects [42]. Our model addresses this complexity by incorporating a variety of variables, which is reflected in the overall improvement in accuracy compared to simpler models.

In line with the literature review, this research also highlights the need to address the problem of student retention from a holistic perspective. While academic performance and attendance are vital factors, our error analysis revealed that they are not unique indicators of the probability of dropping out [43]. This coincides with the existing literature that warns about the complexity of the factors influencing students’ decision making.

Our research focuses on improving student retention and addresses how technological innovation can contribute to sustainability in higher education. Early identification and personalized intervention allow institutions to allocate resources more efficiently, supporting those who need it and reducing spending on blanket initiatives. This is especially relevant in the current context, where the optimization of resources is essential to guarantee the long-term sustainability of educational institutions [18]. The successful implementation of our student retention prediction model has demonstrated the effectiveness of incorporating ML techniques in improving student retention in higher education institutions. Our results support the importance of approaching retention from a multifaceted perspective and provide a promising path for sustainability in higher education.
The findings of our study contribute significantly to the sustainability education goals defined in our research. The results are not only limited to improving student retention but also reveal how applying this definition enriches the educational experience at our university. Students who participated in interventions based on this definition demonstrated a more significant commitment to sustainability issues, a positive change in their attitudes towards collective responsibility, and a more profound development of the skills necessary to address sustainability challenges. These results not only support the relevance of our definition of sustainability education in the context of student retention but also suggest that this definition significantly impacts shaping conscious and active citizens who can address sustainability challenges in society.

Applying ML/AI techniques to improve student retention aligns consistently with the education principles for sustainability. The study has shown that the accurate prediction of student retention and the effective implementation of intervention strategies based on ML/AI can positively impact both the individual student experience and the operational sustainability of the educational institution. By retaining more students through graduation, institutions can optimize their investment in education while improving their financial sustainability. This supports the notion that investing in student retention is an investment in sustainability.

In addition, it is observed that implementing the model leads to higher retention, fosters a positive change in students’ attitudes toward collective responsibility, and promotes the development of essential skills to address sustainable challenges. These results indicate that effective student retention can be a vehicle for forming citizens more aware and committed to sustainability, which goes beyond institutional benefits and contributes to a more solid and equitable educational landscape.

5. Conclusions

The research confirms that implementing the ML-based student retention prediction model has had a significant impact. The results demonstrated an increase of 6.875% in the retention rate, highlighting the effectiveness of the predictions in the early identification of at-risk students and the consequent improvement in retention. The error analysis revealed the complexity of the factors that influence student retention. Although academic performance and attendance remain valuable indicators, they are not unique predictors of dropout. This highlights the importance of a holistic approach and the need to consider socio-emotional and contextual factors in future implementations.

Regarding the research question, it focused on the potential of ML techniques to improve student retention. The results validate this question by demonstrating that the model accurately identifies students at risk, supporting the importance of technological innovation in higher education. In addition to improving retention, our approach has implications for educational sustainability. Early identification and personalized intervention allow for the optimization of resources and the effective allocation of support to needy students. This supports financial sustainability and maximizing retention efforts.

This work has made progress in the accuracy of predicting student retention compared to traditional approaches. By incorporating a diverse set of ML variables and techniques, we have improved the ability to identify at-risk students, overcoming the limitations of simpler models. Our model addresses the interconnection between academic, socio-emotional, and contextual aspects, which enriches understanding and decision making. Our research focuses on retention improvement and presents a sustainability perspective. The optimization of resources and the efficient allocation of support through personalized interventions demonstrate how technology can contribute to sustainability in the educational field.

Comparing the results obtained in this work with the existing literature on student retention and educational innovation, we can highlight several ways our study has had a significant impact. Our research overcomes the limitations of conventional models by offering a substantial improvement in prediction accuracy. This contributes to the development of more effective strategies to address student retention. By considering the interplay
of multiple factors in predicting retention, our approach aligns with the call for a holistic assessment of students. This enriches decision making and attention to individual needs. Implementing ML techniques demonstrates technological innovation’s applicability in the educational field. This integration contributes to updating and modernizing traditional educational practices.

Financial factors play a significant role in student retention challenges in higher education. Issues such as educational affordability, student debt burden, and financial aid availability are vital issues that can influence students’ decisions to continue or drop out of school. Although we cannot provide specific financial data due to the private nature of our institution, we recognize the importance of these factors in the broader context of student retention.

Future work proposes exploring socio-emotional factors, such as psychological well-being and social adaptation, to capture a more complete vision of the student’s situation. Extending the analysis to multiple academic years to understand the long-term impact of interventions and model improvements. In addition, continue to research and address bias in the model to ensure it is fair and accurate for all student groups.

Author Contributions: Conceptualization, W.V.-C.; methodology, J.G.; software, J.G.; validation, S.R.-T.; formal analysis, W.V.-C.; investigation, S.R.-T.; data curation, W.V.-C. and J.G.; writing, original draft preparation, S.R.-T.; writing, review and editing, W.V.-C.; visualization, S.R.-T.; supervision, W.V.-C. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: This work does not require authorization for research in humans. This is considered under the Helsinki Declaration of the Medical Association, where it is established that “Medical research is subject to ethical standards that promote respect for all human beings and protect their health and individual rights. Some research populations are vulnerable and need special protection. The needs of the economically and medically disadvantaged must be recognized. Special attention should also be paid to those who cannot give or refuse consent on their own, those who may give consent under duress, those who will not personally benefit from the research, and those who have research combined with care medical”. According to what is mentioned in the statement, our work is not a medical investigation. On the contrary, what is evaluated is the efficiency and accuracy of the software integrated into an LMS for the improvement of learning through the identification of patterns in grades or patterns that do not invade the privacy of people’s data. Although, for its evaluation, a segment of the population is used, the information collected is clearly in the design and adjustment of software that does not use methods or devices that can be invasive for humans or animals. In addition, by the “REGULATION OF RESEARCH ETHICS COMMITTEES IN HUMAN BEINGS” of Ecuador: (1). Ministerial Agreement 4889 (2). Official Gazette Supplement 279 of 1 July 2014. (3). Status: Current. It establishes “COMMITTEES OF ETHICS FOR RESEARCH IN HUMAN BEINGS (CEISH)” in Art. 4.—Definition.—The Ethics Committees for Research in Human Beings (CEISH) are bodies linked to a public or private institution, responsible for carrying out the ethical evaluation, approving research involving human beings or using biological samples, and ensuring the evaluation and monitoring of clinical studies during their development. Every clinical trial carried out in the country, before beginning its execution, must be evaluated by a CEISH approved by the National Health Authority. For this reason, based on those above, my institution establishes that it cannot issue any certificate as it is not medical research work, considering that the study’s objective is the design of software as an assistant for an academic environment without the publication of data that invades the privacy of the participants. Attached is the official record of Ecuador’s HUMAN RESEARCH ETHICS COMMITTEE (CEISH).

Informed Consent Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.
References


10. Forney, A.; Mueller, S. Causal Inference in AI Education: A Primer. *J. Causal Inference* 2022, 10, 1–33. [CrossRef]


32. Duan, X.; Fan, M.; Fan, S. Comprehensive Evaluation of Structural Variant Genotyping Methods Based on Long-Read Sequencing Data. *BMC Genomics* 2022, 23, 324. [CrossRef]


38. Arora, A.; Arora, A. Generative Adversarial Networks and Synthetic Patient Data: Current Challenges and Future Perspectives. *Future Healthc. J.* 2022, 9, 190–193. [CrossRef]


41. Karimian, G.; Petelos, E.; Evers, S.M.A.A. The Ethical Issues of the Application of Artificial Intelligence in Healthcare: A Systematic Scoping Review. *AI Ethics* 2022, 2, 539–551. [CrossRef]

42. Chicco, D.; Jurman, G. The Advantages of the Matthews Correlation Coefficient (MCC) over F1 Score and Accuracy in Binary Classification Evaluation. *BMC Genomics* 2020, 21, 6. [CrossRef]


Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.