sustainability

MDPI

Article

Developing a Model to Predict Self-Reported Student Performance during Online Education Based on the Acoustic Environment

Virginia Puyana-Romero [1,*], Cesar Marcelo Larrea-Álvarez [2], Angela María Díaz-Márquez [3], Ricardo Hernández-Molina [4] and Giuseppe Ciaburro [5]

1 Department of Sound and Acoustic Engineering, Faculty of Engineering and Applied Sciences, Universidad de Las Américas (UDLA), Quito 170503, Ecuador
2 Faculty of Medical Sciences, Medical Career, Universidad de Especialidades Espíritu Santo, Guayaquil 092301, Ecuador; cmla88@hotmail.com
3 Innovation Specialist in Higher Education, Information Intelligence Directorate, Universidad de Las Américas (UDLA), Quito 170503, Ecuador; angela.diaz@udla.edu.ec
4 Laboratory of Acoustic Engineering, Universidad de Cádiz, 11510 Puerto Real, Spain; ricardo.hernandez@uca.es
5 Department of Architecture and Industrial Design, Università degli Studi della Campania Luigi Vanvitelli, Borgo San Lorenzo, 81031 Aversa, Italy; giuseppe.ciaburro@unicampania.it
* Correspondence: virginia.puyana.romero@udla.edu.ec or virginiapuyana@yahoo.es

check for updates

Citation: Puyana-Romero, V.; Larrea-Álvarez, C.M.; Díaz-Márquez, A.M.; Hernández-Molina, R.; Ciaburro, G. Developing a Model to Predict Self-Reported Student Performance during Online Education Based on the Acoustic Environment. *Sustainability* **2024**, *16*, 4411. https://doi.org/10.3390/su16114411

Academic Editor: Hao-Chiang Koong Lin

Received: 10 April 2024
Revised: 26 April 2024
Accepted: 8 May 2024
Published: 23 May 2024

Correction Statement: This article has been republished with a minor change. The change does not affect the scientific content of the article and further details are available within the backmatter of the website version of this article.

Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

Abstract: In recent years, great developments in online university education have been observed, favored by advances in ICT. There are numerous studies on the perception of academic performance in online classes, influenced by aspects of a very diverse nature; however, the acoustic environment of students at home, which can certainly affect the performance of academic activities, has barely been evaluated. This study assesses the influence of the home acoustic environment on students' self-reported academic performance. This assessment is performed by calculating prediction models using the Recursive Feature Elimination method with 40 initial features and the following classifiers: Random Forest, Gradient Boosting, and Support Vector Machine. The optimal number of predictors and their relative importance were also evaluated. The performance of the models was assessed by metrics such as the accuracy and the area under the receiver operating characteristic curve (ROC_AUC-score). The model with the smallest optimal number of features (with 14 predictors, 9 of them about the perceived acoustic environment) and the best performance achieves an accuracy of 0.7794; furthermore, the maximum difference for the same algorithm between using 33 and 14 predictors is 0.03. Consequently, for simplicity and the ease of interpretation, models with a reduced number of variables are preferred.

Keywords: online learning; domestic soundscape; self-reported academic performance; noise sources; machine learning

1. Introduction

It seems inevitable that the changing trends in educational systems will continue toward a situation where, due to the increasing refinement of Information and Communication Technology (ICT), online courses will become part and parcel of the learning experience. Regardless of their limitations, Massive Open Online Courses (MOOCs) have considerable potential for teaching [1] and evident benefits for students [2]. Furthermore, specific research patterns have been identified around MOOCs. Although a shift towards a more critical discourse is noticeable, positive perspectives are prevalent over negative ones. Despite researchers' preference for theoretical or conceptual studies, the overall benefit of viewing MOOC research through such lenses is limited [3]. With a growing number of students opting for online courses, understanding their relative effectiveness through experimental studies has become progressively imperative for the improvement of

education. For this reason, researchers allot serious efforts to evaluating the effects of the online methodology, either based on student satisfaction or student performance and its potential explicative factors [4–11]. So, because the COVID-19 lockdown forced a situation where activities significantly changed for a while, it offered a unique scenario to assess the status, challenges, and the implications of the online teaching approach as the unique alternative. Researchers seized this opportunity and gathered considerable data to further characterize student experience with online courses.

If the purpose is to assess student performance, we should consider the conditions of their working environment, which, in general, is a decisive aspect for improving productivity [12,13]. It is fair to say that a critical factor determining the quality of the work environment is the level of interference and background noise. Findings on this issue indicate that undertaking academic work becomes increasingly demanding as environmental noise exceeds the permissible limit, which, in turn, undermines student performance [14]. Also, the negative impact of noise on learning activities has been reported in physiological terms, from dizziness to feelings of discomfort [15]. Excessive noise has been identified as a significant disruptor, compromising students' ability to comprehend information delivered by instructors. In traditional face-to-face classes, background noise interference often leads to the reception of unclear and distorted messages by students [16–18]. These acoustic challenges pose obstacles to effective learning, particularly impacting individuals dealing with adaptation issues, hearing difficulties, those whose native language differs from the residence location, or those engaged in the process of learning a second language [18,19]. The implications of these acoustic disruptions underscores the need for a comprehensive understanding of the home learning environment and its potential effects on students' educational experiences.

The analysis of online teaching effectiveness based on comparisons with face-to-face (F2F) modalities offers conclusions that, on the one hand, reject the balance of validity between both paradigms and, on the other hand, vindicate the merits of online courses over the F2F method. First, among the critical views, some arguments consist, for example, in the lack of replication of critical conditions present only in F2F classrooms [15], in the disparity of content of discussions, which commonly leads to a clearer understanding by F2F students [12], and in the notion of the "McDonaldization" of the learning process [20]. The second stand, which carefully contemplates online education as a feasible substitute for classical classroom education, usually refers to the appealing qualities of distance learning for students, e.g., the flexibility concerning time and location, a higher diversity of programs and institutions [21], and even the provision of a more comfortable environment for introvert students [22]. This whole body of results relies heavily on standard performance evaluation data, usually collected through varied criteria, which is probably the reason for such indeterminate conclusions. In this context, in the study carried out by Discroll et al. [23], the differences in student performance and satisfaction between online and face-to-face classroom settings were evaluated. They set a course design where a "deliberate effort was made to keep the two types of classes as similar as possible", allowing them to test such potential differences employing the "pedagogical practices" applied to each type of class in a more controlled manner. Although they found differences in the performance between groups, they explained them away by a "selection effect", expressed by the self-reported estimated student Grade Point Average (GPA), and caused by a bias of higher presence of higher GPA students on F2F courses. The latter is manifested in the behavior of their Ordinary Least Square regression (OLS). This way, they suggest that it is this bias that "creates the appearance of the online classroom being a less effective learning environment", and that there is no inherent deficiency in the online approach. By the same token, Paul and Jefferson [24] evaluated the influence that the "instructional medium" (online vs. F2F) has on student performance. Using a substantial amount of data covering an 8-year Environmental Science class, they compared the final GPA grades of both modalities and ended up finding no significant differences between them. Thus, they suggest that the "teaching modality may not matter as much as other factors",

and can accentuate the social relevance of such implications, such as facilitating access and promoting proper education to the general public, through online courses. Further comparative studies have supported these conclusions, indicating that students from online or hybrid classes could be more productive than F2F students [25–27]. Additionally, as we have noted before, during the COVID pandemic, researchers studied the implications that these "special" and new conditions had on students. For instance, Gopal et al. [10] attempted to pinpoint the relevant factors that would inform educators on how to attain, under these unexpected circumstances, a proper satisfaction level from students and, in this way, a higher performance. They suggest that their results "can be used to continually improve and build courses", as well as to implement laws to enhance educational programs.

With this trend of diminishing skepticism towards the online teaching methodology, and with the aid of refined artificial intelligence algorithms, researchers have tried not only to characterize student performance, but also to predict it [28]. Aydoğdu [29] proposed to model student performance in online learning environments, employing a set of predictors that represent subject behavior. Using data from 3518 university students, they implemented and optimized an artificial neural network to find which parameters contribute the most to the highest-accuracy prediction of student performance. They report an effective predictive potential of the model with an achieved accuracy of 80.47%. Finally, they offer future predictive modeling based on other possible explicative variables, such as student "instant browser data". Analogously, Alshabandar et al. [30] applied predictive models on student performance to try to identify the factors that influence learning achievements, but they contemplated the examination of performance trajectories. The intention was to help educators to be able to monitor students' learning curves. Their model predicted the performance with three class labels "success", "fail" and "withdrew". The final model showed practical and accurate results (based on the applied metrics) and revealed how tracking student performance, i.e., considering temporal features, would improve the subsequent assessments of grades. Finally, it is worth mentioning that the study conducted by Segura et al. [31], where, in an attempt to help identify at-risk students in the context of MOOCs, a high-accuracy model to predict potential dropouts was constructed. By adopting five popular algorithms, Artificial Neural Network (ANN), Support Vector Machine (SVM), K-Nearest Neighbors (KNN), decision trees, and logistic regression, they show that, from an algorithmic perspective, the power of deep learning predominates over the others and that among the more traditional methods, SVM is less prone to overfitting. Their model is proposed to educators as a possible personalized intervention to prevent student dropout, based on the prediction of dropout probabilities.

Research must assess various frameworks to ensure the long-term sustainability of online education beyond the COVID-19 lockdown [32–34]. This evaluation should adopt a holistic approach, which considers both the individual circumstances of students and the casuistics of new online academic modalities. However, since the confinement represented a potential source of knowledge to establish the foundation for future research, before undertaking this holistic approach, it is important to understand in isolation the different aspects that affected the student learning process during the COVID-19 confinement.

While numerous research studies have extensively explored a wide range of factors influencing academic performance in online education, the impact of the acoustic environment has been largely overlooked. The purpose of this study was to state the basis for developing a methodology that considers acoustic factors in the perceived academic performance during online education. For that, models to predict the perceived academic performance, based on acoustic variables, were calculated using three machine learning algorithms within the Recursive Feature Elimination method: Random Forest, Gradient Boosting, and SVM. The relative importance of the variables was calculated to avoid collinearity and to improve the efficiency of the models.

- The following research questions were addressed and answered in the study:
- Can a predictive model perform well with a reduced number of predictors, most of them acoustics?

- Do the variables selected for the predictive model avoid repeating acoustic information?
- Are acoustic variables important for the prediction of perceived academic performance?

2. Materials and Methods

To assess the impact of the acoustic environment on the students' perception of their academic performance, an online survey was conducted (Section 2.1). Using the collected information, a dataset was generated, primarily focusing on variables associated with their home acoustic environment. Spearman correlation coefficients were computed to determine the presence of linear associations between variables (Section 2.2). Given the imbalanced nature of the dependent variable, the Random Forest Classifier algorithm was applied to various imbalanced and resampled dataset scenarios to determine the most suitable dataset for the subsequent feature selection stage (Section 2.3). Based on the outcomes, and to obtain the most efficient number of relevant variables for the conformation of the models, the Recursive Feature Elimination method was applied with three different algorithms: Random Forest, Gradient Boosting, and Support Vector Machine (Section 2.4).

Subsequently, the most relevant variables were selected (Section 2.5), and the predictive models were calculated. Diverse metrics were used to evaluate the performance of the models. A comparison was conducted between the models calculated with the optimal number of variables for each algorithm, and the models calculated with the reduced number of variables (smallest number of variables of all) to assess the efficiency of both types of the predictive models (Section 2.6).

2.1. Online Survey

Conducted over 20 days in January and February 2021, the online survey was accessible exclusively to the students of the Universidad de Las Américas in Quito, Ecuador. By employing Microsoft Forms, the questionnaire design (written in Ecuadorian) ensured that only those within the university's inner network could respond. Participants received invitation emails containing a link granting them access to the survey. Before commencement, the questionnaire explicitly stated that completing it implied the student's authorization for the use of provided data in research. The survey garnered responses from 2477 participants. To safeguard confidentiality, a numerical code was assigned to each participant.

In terms of content, the survey encompassed inquiries related to personal information (age, gender, and semester), the domestic acoustic environment (identifying noise sources and levels), and noise interference during academic activities [11]. According to the type of activity, they were categorized as synchronous or autonomous, and the responses were classified on Likert scales ranging from 5 to 7 points. Table A1 in Appendix A provides a summary of the survey questions regarding the domestic acoustic environment, and Table A2 shows the GPA of each semester, together with their corresponding acronyms for reference in the study. The semester GPA in which the survey was conducted was called GPA 202110.

2.2. Dataset

Our dataset consists of 40 variables, 33 of which represent the perception of intensity and the recurrence of noise interference. The rest stand for perception of teaching quality ACADEMIC_Q (3 variables), GPA ratings (3 variables), and the experience of internet connection problems, "INTER_AUDIO" (1 variable). To simplify the distribution of responses and better understand the results, the dependent variable, which reported the perception of the student's performance ("ACAD_PERFO"), was codified as "GOOD" (scores > 4) and "BAD" (scores $\leq$ 4) for the study.

To evaluate the relationship between the variables, and considering that most of them were ordinal [35,36], Spearman's correlation coefficients were calculated. These coefficients measure the strength and direction of the relationship between two variables. In this case, the Spearman coefficient was used due to the ordinal nature of the variables. Unlike

the Pearson correlation coefficient, which is used for continuous variables, the Spearman coefficient is appropriate for ordinal data.

2.3. Evaluating the Effects of Dependent Variables' Imbalanced Distribution

To gain better insight and information about the data, an exploratory analysis on the imbalanced nature [37–40] of the dependent variable ("ACAD_PERFO") was carried out. This would determine which dataset to use in the feature selection process.

The calculations were performed with Python 3.9.12, using the Sklearn [41] and Imblearn [42] libraries (from packages Scikit-learn v1.0.2) and Imblearn v0.10.1) respectively), which feature the required algorithms for the exploratory analysis.

The imbalanced nature of the dependent variable was assessed by carrying out the following processes: (a) arranging different dataset scenarios; (b) applying Random Forest to each scenario; and (c) calculating the area under the receiver operating characteristic (ROC) curve (AUC-score).

(a) The following dataset scenarios were generated:
   - (1) Data with its original class distribution and the classifier considering the equal (=1) weights for each class.
   - (2) Random over-sampling in a 1:1 ratio;
     - (2.1) only over-sampling the training dataset;
     - (2.2) over-sampling all the data.
   - (3) Random under-sampling in a 1:1 ratio;
     - (3.1) only under-sampling the training dataset;
     - (3.2) under-sampling all the data.
   - (4) Data with its original distribution and the classifier considering adjusted ("balanced") weights for each class.
   - (5) Combination of over- and under-sampling to obtain a less imbalanced dataset;
     - (5.1) only on the training dataset.
     - (5.2) on all the data.

(b) A Random Forest algorithm on the previously imbalanced and resampled dataset scenarios was applied to explore which dataset to use in the feature selection process: the original class distribution or the resampled data after the 'handling-imbalance' procedure. To each scenario, 5 repetitions of 10 iterations were applied. For each iteration: (1) the data were split into train and test sets using the function cross_val_score(). Only for scenarios 4, 5.1, and 5.2, the parameter class weight was set as "balanced", which automatically adjusts weights inversely proportional to class frequencies. (2) A Random Forest classifier was fitted to each dataset (according to the scenario configuration, the train, the test sets, or both could maintain their original weights, or be under-sampled, over-sampled, or "balanced").

(c) With enough imbalance, some relevant metrics assessing the model performance could be misleading given the bias towards the majority class during the training stage [43]. For the current assessment, the AUC-score metric was calculated which is generally considered to perform well with imbalanced data [44]. Consequently, to obtain an objective insight, the predictions were evaluated, calculating the mean AUC score for the test set of each scenario.

2.4. Recursive Feature Elimination

The technique of recursive feature elimination (RFE) is widely used in data analysis and is employed to detect and eliminate the irrelevant or redundant features in a dataset. This process contributes to improving both the accuracy and efficiency of the machine learning model by reducing the number of redundant features. The RFE is adaptable to any type of algorithm, rendering it a versatile method that is applicable across various scenarios. The Recurse RFE method is classified by some authors as a wrapper method [45]. Some

other sources consider it a wrapper-type feature selection method that also uses filter-based feature selection internally [46]. Wrapper and filter methods have some differences, but the main one lies in their dependency on the learning algorithm. To apply a filter method is not necessary to know a priori which learning algorithm is going to be used; they score each feature and select the ones with the largest or smallest score. In contrast, Wrapper methods, however, use an algorithm at the core of the method that makes an optimal choice at each stage, ranking features by importance and fitting the provisional model. This selection process is performed until the model is completely fitted [47].

Wrapper methods tend to overfit the models, as they conduct feature selection with several sets of feature combinations. To avoid this, the dataset is split into train and test subsets. RFE starts with all the features of the training subset and iteratively removes features until the desired number remains. The selection algorithm used in the core of the RFE method ranks the given features in order of importance. Different stages conduct iteratively the previous process until the model is fit. The selection algorithm facilitates the computation of significant scores and can be, for instance, a decision tree. However, the algorithm employed within the RFE does not necessarily have to be applied the same to select the features; alternative algorithms can be integrated [46,47].

In the previous step (evaluating the effects of dependent variables' imbalanced distribution), once the Random Forest algorithm was calculated for each scenario, the best-performed dataset was selected to avoid the effect of imbalanced data. Subsequently, RFE was conducted on this dataset to select the optimal number of features that predict (or classify) the target variable. The objective was to reduce the dependencies and collinearity that may exist. For that, a two-step process was performed: (1) searching for the optimal number of features, and (2) selecting these features based on their hierarchy of importance. The procedure applying the Recuse Feature Elimination was carried out with Python 3.9.12, using the Scikit-learn (v1.0.2) packages and the RFE() function [41]. The input parameters included the estimator and the number of features. The RFE() function uses a specific classifier (set in as an estimator) to assign weights to features. Then, to select the appropriate variables, it applies a recursive process that considers and evaluates smaller and smaller sets of variables by removing features recursively; if one of the collinear features is removed, the coefficient or feature importance of the remaining feature(s) increases.

The following configuration was applied to calculate the models: given a classification algorithm and a specific number of features to select, the RFE model using repeated stratified k-fold cross-validation was assessed, with three repeats and 10-folds. The cross-validation guarantees that each time a feature is removed, the performance of the resulting model is calculated to assess if the removal of the feature was beneficial, providing a stable and robust set of selected predictors. The function cross_val_score() was applied, with a splitting strategy determined by the function RepeatedStratifiedKFold(), and the AUC metric as estimator scoring. The score was obtained by fitting the RFE and subsequently using the result in a given classifier. The attained mean AUC for each classified algorithm was used to tune the optimal number of input features (among 40). The RFE function was carried out for each of the following classifier algorithms: Random Forest (with and without "balanced" classes), Gradient Boosting, and SVM (Linear Kernel).

Some machine learning algorithms are sensitive to feature scaling, while others are virtually invariant to it. For example, Gradient Descent Base algorithms require data to be scaled to help data to converge more quickly towards the minima [48]. Similarly, the data is normally scaled to apply Distance-Based Algorithms (e.g., KNN, K-means, and SVM), so that all the features contribute equally to the result [49,50]. However, Tree-Based Algorithms, such as Decision Trees, Random Forests, and Gradient Boosting are not sensitive to the magnitude of variables. The split of nodes by a feature is not influenced by other features, thus, standardization is not needed before fitting these kinds of models [48,51]. For this reason, the above analysis was initially performed with standardized data using all algorithms. Subsequently, the process was repeated only with the tree-based algorithms, but this time using non-standardized data to evaluate possible differences.

2.5. Features Selection and Features Importance

In the previous step, different algorithms were tested within the RFE method. For each algorithm, the number of selected features that achieved the best performance was obtained. Subsequently, the best AUC score and the optimal number of features were the criteria adopted to select an algorithm, with an optimal number of selected features (n_f) to be used in this step.

For the feature selection process, the strategy was to: (1) Instantiate a RFE model, and set it to n_f features with a classifier as an estimator (as previously highlighted, the classificatory algorithm that scores the variables does not have to be the same as the one used to fit the model and calculate the relative importance of the variables). Subsequently, fit the RFE model with the whole dataset to obtain the most relevant features. (2) Obtain a reduced dataset, consisting of these n_f selected columns and the dependent variable. The algorithm with better performance was used to obtain the impurity-based importance of each variable. (3) Finally, for the same reduced dataset, 3 repetitions of a 10-fold cross-validation process were conducted to evaluate the selected variables with the best-performed classifier, and its performance was evaluated using the AUC metric. Again, the method involved was the cross_val_score() function (which selects a more robust set of predictors for the final model), with a splitting strategy specified with RepeatedStratifiedKFold(). More information about the implementation of the feature selection process calculated in Python 3.9.12 can be found in the work presented in [41]. In this way, all the data was used and internally partitioned into training and test sets to compute the corresponding AUC metric. Thus, the mean AUC metric of 30 results (calculated from 3 repetitions $\times$ 10 iterations) was obtained from testing the previously selected features in a classifier.

2.6. Implementation of Classifiers

With the reduced dataset composed of a specific number of variables with the best AUC score, three classifiers were implemented in R 4.2.3 software using the packages caret (v6.0-94), e1071 (v1.7.14), and pROC (v1.18.5): Random Forest, SVM, and Gradient Boosting. For each method: (1) the data was split into training (80%) and test (20%) sets; (2) the relevant hyper-parameters were tuned with a cross-validation 10-fold grid search; (3) the models were trained with the tuned parameters; (4) the final classifier was evaluated; and (5) the area under the ROC curve was calculated.

According to the literature [48–51] and the results obtained, the data used to calculate the SVM models were standardized, while for the Random Forest and Gradient Boosting classifiers were not.

The data partition was performed with the caret function createDataPartition(), which assures that for both the training and test datasets, the dependent variable retains its class imbalance. For hyper-parameter optimization, the caret train() function tuned the respective values using a grid search cross-validation control strategy set through a caret trainControl() object. The train() method currently possesses the appropriate algorithms for our purposes, and for each classifier, the criteria included calculations started with a set of values for a specific hyper-parameter, and subsequently, this set was tested. Then, if the best-performing value was not the largest in the inspected group, that one was picked. Otherwise, another set of hyper-parameters was defined to be evaluated, including the previously selected one, and subsequently, it was tested. After that, the classifier was fitted with three repetitions of a 10-fold cross-validation grid search, using the caret train() method and the tuned hyper-parameters. Next, employing the generic predict() function, the fitted model was evaluated with the test dataset. From this, a confusion-matrix analysis was carried out, applying the caret function confusionMatrix(), which allowed us to compute, among others, the accuracy, McNemar's test *p*-value, the Positive Predictive Value, the Negative Predictive Value, and the ROC-AUC [52]. Finally, because the assessment was focused on the AUC score, the ROC plot was depicted as an additional way to show the model's ability to predict classes correctly. The latter was performed using the pROC roc() function.

The previous metrics are based on the parameters used to build a Confusion matrix [53,54] which displays the number of instances produced by the model on the test data: true positives (TP) occur when the model accurately predicts a positive data point; true negatives (TN) occur when the model accurately predicts a negative data point; false positives (FP) occur when the model predicts a positive data point incorrectly and false positives (FN) occur when the model predicts a negative data point incorrectly. The Accuracy [40] is the metric that shows the proportion of the correctly predicted cases out of the total cases (TP + TN)/Total. McNemar's test *p*-value is a non-parametric test that, in the context of machine learning, is used to compare the predictive accuracy of two models [55]. The null hypothesis involves that none of the two models performs better than the other. For the calculations, one model is the implemented model and the other is the baseline model. A baseline model, or dummy model, is a classifier model that makes predictions without attempting to find patterns in the data. The Positive Predictive Value [56], also called precision, is a metric that reflects, out of the total predicted positive cases, how many belong to the actual positive class, TP/(TP + FP). Consequently, the Negative Predictive Value indicates, out of the total predicted negative cases, how many belong to the actual negative class, TN/(TN + FN).

The Confusion matrix previously described and the ROC curve are very useful visual tools to assess and compare the performance of the models. The ROC curve visually depicts how well a classification model performs across different thresholds for classifying data [57–59]. It is plotted between the true positive rate (TPR, also called recall) on the *y*-axis against the false positive rate (FPR) on the *x*-axis at various threshold settings. The AUC score is the area under this ROC curve, meaning that the resulting score represents, in broad terms, the model's ability to predict classes correctly. It performs well with Imbalanced data. As a reference, an AUC score of 0.5 indicates that the ROC curve falls on the diagonal (i.e., 45-degree line) and hence suggests that the diagnostic test has no discriminatory ability. Although the literature establishes a wide variety of intervals to evaluate the goodness of this metric [60], as a reference, the following ranges have been considered in this study to interpret its discrimination ability [61]: 0.5 = random; [0.5, 0.59] = poor; [0.6, 69] = moderate; [0.70, 0.79] = good; [0.8, 0.89] = very good and, [0.9, 0.99] = excellent.

Once the metrics were calculated, the performance of the models with the optimal number of variables was compared. To guarantee the importance of the acoustic environment, a model only with the selected acoustic variables and the best-performed algorithm was built, and its performance was compared to the previously developed model.

3. Results and Discussion

3.1. Descriptive Statistics

Spearman correlations have been calculated between the 41 variables involved in the study. To simplify the results, only the correlations between some variables were shown. Spearman correlation coefficients of the variables "ACAD_PERFO" (target variable), "GPA_202010", "GPA_202020", "GPA_202110", "INT_SYNC" and "INT_AUTO" are shown in the upper panel of Figure 1. The lower panel shows the scatter plots between the variables, the diagonal, and the density graphs.

The density graphs show that the GPA distribution is similar for the three semesters considered in the study, especially taking into account that most of the 0 values correspond to students who had started their studies in the semester in which the survey was conducted (202110), and had no marks in the previous semesters (202020 and 202010). These graphs also show that most students have neutral (score = 4) or positive self-perceptions of their academic performance (scores $\geq$ 4). Furthermore, the majority of the students consider that domestic noise has interfered with their academic synchronous and autonomous activities (scores > 1), although there is a representative group of them that believe that noise did not bother them while conducting their academic activities (scores = 1).

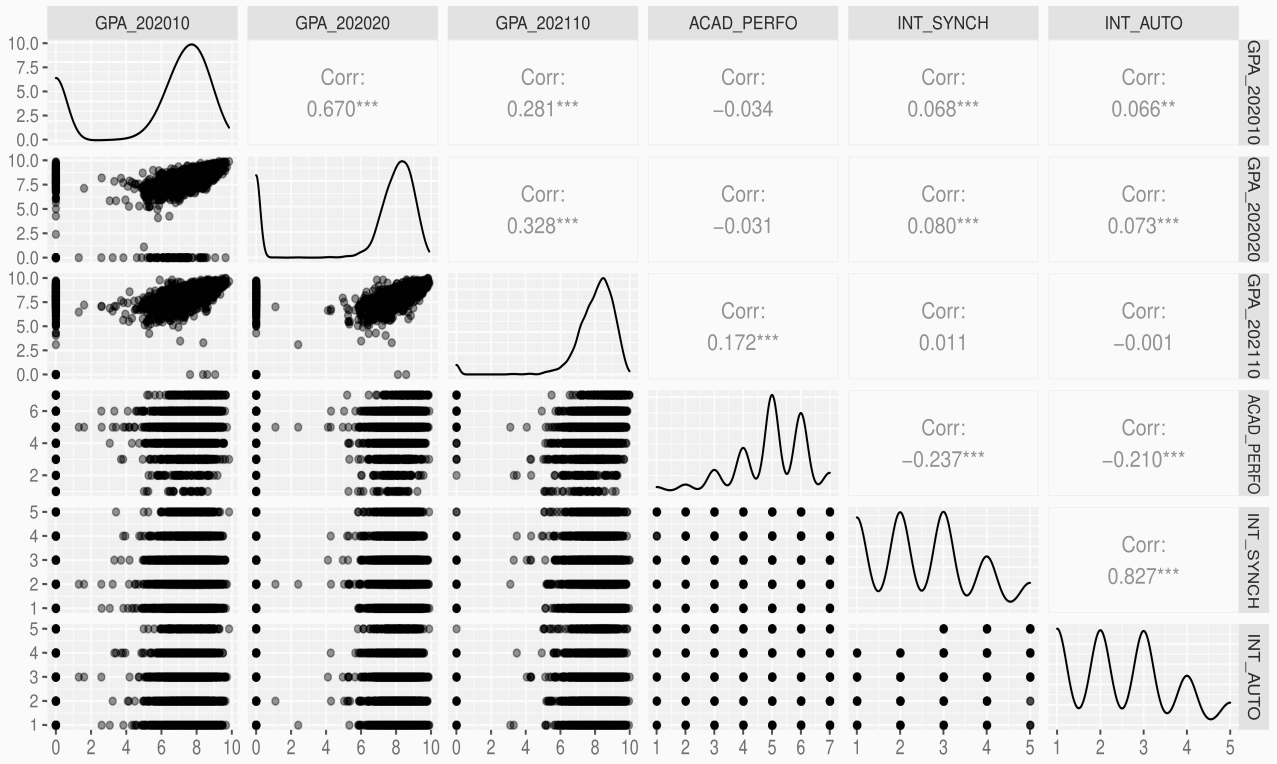

Figure 1. Above the diagonal, the Spearman correlation coefficient between the variables "ACAD_PERFO" (target variable), "GPA_202010", "GPA_202020", "GPA_202110", "INT_SYNCH" and "INT_AUTO". Significant positive or negative correlations are marked with ** $p < 0.01$, and *** $p < 0.001$. The diagonal shows the density graphs. Below the diagonal, scatter plots between the variables are depicted. The acronyms correspond to the responses given to different questions in the survey (see more information in Appendix A).

According to Figure 1, there is a positive correlation with low strength [62] (r = 0.172) between the GPA of the semester in which the survey was taken ("GPA_202110") and the perception of student performance ("ACAD_PERFO"), although with a degree of significance *p*-value less than 0.001, which makes it statistically significant. The perception of the noise interference in synchronous ("INT_SYNCH") and autonomous ("INT_AUTO") activities shows also a significant but negative correlation strength, with the correlation coefficients equal to −0.237 and −0.210, respectively.

3.2. Evaluating the Effects of Dependent Variables' Imbalanced Distribution

The target variable ("ACAD_PERFO") is slightly imbalanced, with 28.2% for the minority class. The minority class for the confusion matrix analysis is the case labeled as "BAD" (Figure 2). The imbalanced ratio (IR), as defined in the work presented in [37], is IR = minority class/majority class = 2.55.

Due to the imbalanced nature of the dependent variable, different datasets and Random Forest configurations were arranged to determine which dataset could better fit the purposes of the study, as explained in Section 2. The ROC-AUC score calculated on the test set was used to evaluate the predictions.

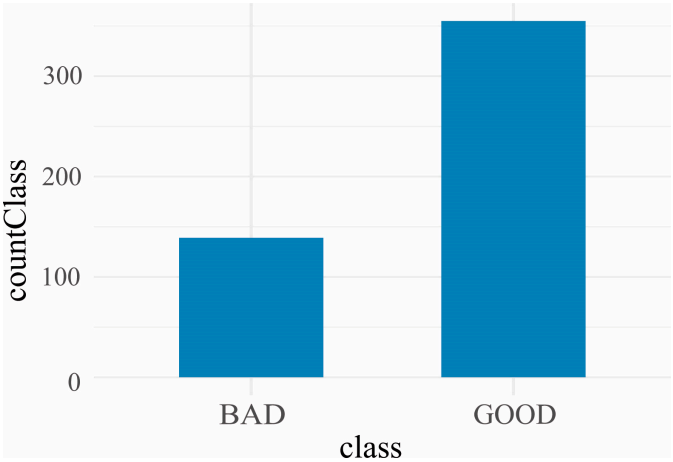

Figure 2. Target variable distribution. "GOOD" = 71.8% (1779 counts) and "BAD" = 28.2% (698 counts).

Figure 3 shows the distribution of the AUC score of the models calculated for the different scenarios using boxplots. Scenarios 2.1, 3.1, and 5.1 have a smaller variability in the AUC score, but also smaller medians. Table 1 shows the Mean AUC score and the standard deviation of the different scenarios analyzed. According to Table 1 and Figure 3, if scenario #1 is considered the control case, it can be seen that none of the "experiments" improve its mean AUC score metric (0.7634), which suggests that there is no reason to resample the data, neither for the Recursive Feature Elimination process, the optimization of hyper-parameters, and for the training of the models.

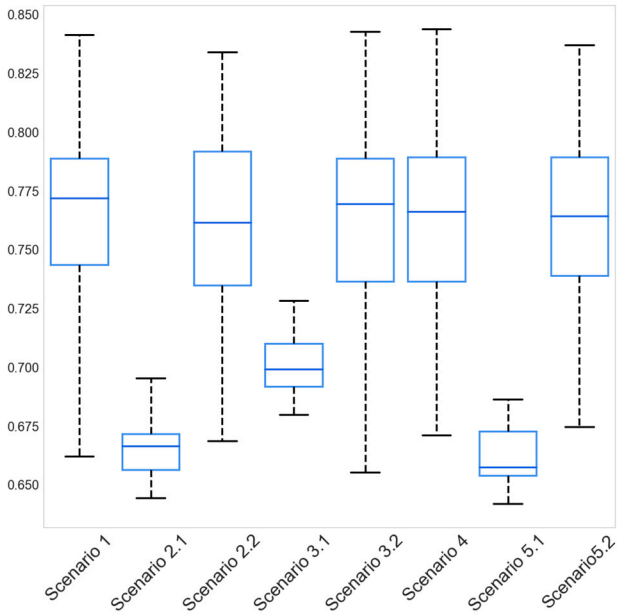

Figure 3. Boxplot of the AUC scores calculated for the different scenarios. From bottom to top, the first blue line of the box shows the first quartile, the second one shows the median value, and the last, the third quartile.

Table 1. Mean AUC-score and standard deviation (Sd) of the different dataset scenarios to evaluate the influence on the results of the imbalance nature of the dependent variable. The AUC score is the area under the ROC (receiver operating characteristic) curve.

Scenario	Mean AUC Score	Sd
Scenario #1	0.7634	0.0381
Scenario #2.1	0.6655	0.0114
Scenario #2.2	0.7602	0.0381
Scenario #3.1	0.7011	0.0116
Scenario #3.2	0.7631	0.0378
Scenario #4	0.7628	0.0377
Scenario #5.1	0.6616	0.0121
Scenario #5.2	0.7612	0.0369

3.3. Recursive Feature Elimination

Number of Features Selection

Given an estimator algorithm, and to obtain the optimal number of features, the RFE method was assessed for the 40 possible number of features, using repeated stratified k-fold cross-validation, with three repeats and 10-folds. The applied function was the cross_val_score(), and the 'AUC' metric was used as estimator scoring.

The attained mean AUC score was considered to compare the performance of the models. Figure 4 shows, at each point, the mean AUC score for each of the 40 possible number of features tested in RFE for each estimator algorithm used: Random Forest with (d) and without balanced classes (a), SVM (c), and Gradient Boosting (b). Table 2 summarizes the Mean AUC score and the standard deviation of the models with the optimal number of variables selected for each algorithm using standardized data. The optimal number of predictor variables to build models with standardized data is 34 for the Random Forest algorithm (with imbalanced data) (mean AUC score = 0.7466, sd = 0.0387), 18 for the Gradient Boosting algorithm (mean AUC score = 0.7466, sd = 0.0387), 32 for Random Forest (with balanced data) (mean AUC score = 0.7451, sd = 0.0383), and 28 for SVM (mean AUC score = 0.7506, sd = 0.0182).

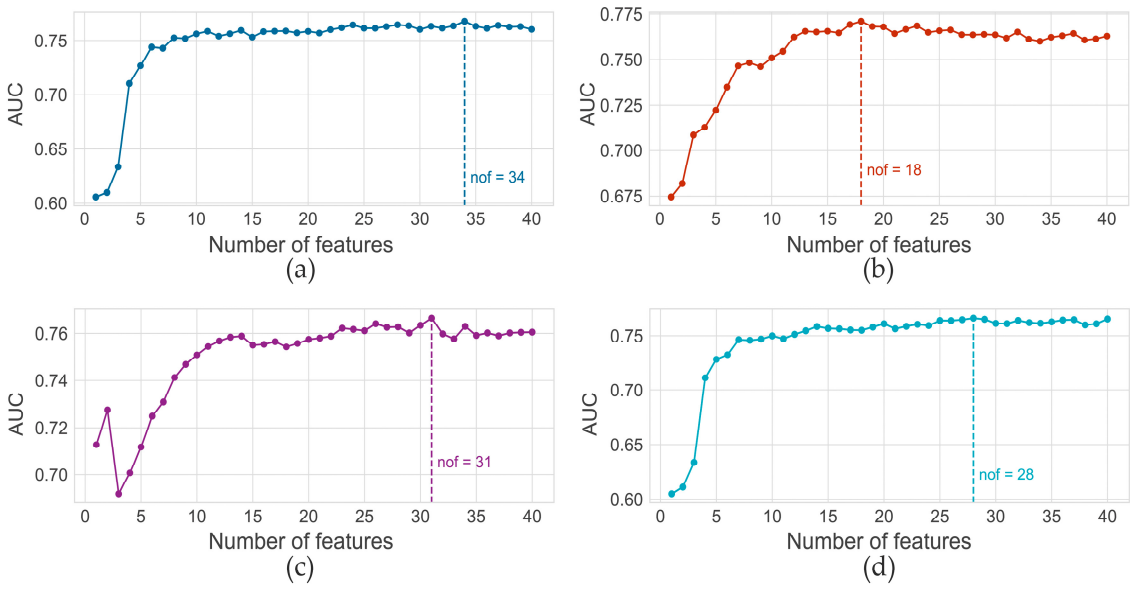

Figure 4. Mean AUC score of each number of variables used in 40 models calculated using the following algorithms (**a**) Random Forest, (**b**) Gradient Boosting, (**c**) Support vector machine (SVM), and (**d**) Random forest, calculated with balanced data. The dotted line indicates the number of features that lead to the model with the highest Mean AUC score. The models were calculated with standardized data.

Table 2. Mean AUC-score and standard deviation (Sd) of the models calculated with four algorithms used within the Recursive Feature Elimination method to evaluate the optimal number of variables of the model (last column). The AUC score is the area under the ROC (receiver operating characteristic) curve.

Algorithm	Mean AUC Score	Sd	Number of Variables
Random Forest	0.7466	0.0387	34
Gradient Boosting	0.7538	0.0225	18
Random Forest (balanced)	0.7451	0.0383	32
SVM (linear)	0.7506	0.0182	28

Some sources indicate that the tree-based algorithms are not sensitive to the scale of the variables [48,51]. Subsequently, the previous calculations were conducted again using the decision tree algorithms of our study (Random Forest and Gradient Boosting), but this time using non-standardized data. Table 3 and Figure 5 summarize the results. The optimal number of predictor variables to build models with non-standardized data is 33 for the Random Forest algorithm (with not balanced data) (mean AUC score = 0.7467, sd = 0.0381), and 14 for the Gradient Boosting algorithm (mean AUC score = 0.7537, sd = 0.0227).

Table 3. Mean AUC score and standard deviation (Sd) of the models calculated using tree-based algorithms in the Recursive Feature Elimination method to evaluate the optimal number of variables, using non-standardized data. The optimal number of variables that lead to the best-performing model is shown in the last column. The AUC score is the area under the ROC (receiver operating characteristic) curve.

Algorithm	Mean AUC Score	Sd	Number of Variables
Random Forest	0.7467	0.0381	33
Gradient Boosting	0.7537	0.0227	14

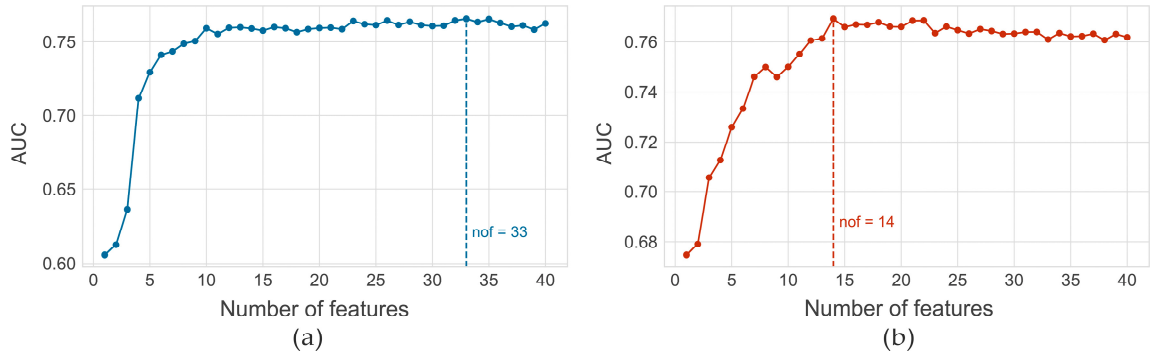

Figure 5. Mean AUC score of each number of variables used in models calculated using the following algorithms, using non-standardized data: (**a**) Random Forest and (**b**) Gradient Boosting. The dotted line indicates the number of features that lead to the model with the highest Mean AUC score.

The comparison between Figure 5 and Table 3 with Figure 4 and Table 2 shows that there was no improvement in the Mean AUC score using standardized data for the tree-based algorithms. Gradient Boosting with standardized data was the algorithm with the highest Mean AUC score (0.7538), with a standard deviation of 0.0225. However, the differences with the model calculated with the same algorithm and non-standardized data were insignificant (dif. Mean AUC score = 0.0001 and dif. Sd = 0.0002, but this model used the lowest number of variables among all (14). For these reasons, for the next step (features selection and features importance), the Gradient Boosting method with non-standardized data was used.

3.4. Features Selection and Features Importance

After calculating the optimal number of features, we wanted to know which 14 features (among 40) should be considered for the calculation of the prediction models. The strategy was: (1) instantiating a RFE model, and setting it to 14 features with a Random Forest Classifier as an estimator. A RFE model was fitted with the whole dataset to obtain the most relevant features. (2) The reduced dataset was selected, consisting of these 14 selected columns and the dependent variable, and fitted with a Gradient Boosting classifier to obtain the impurity-based importance for each variable. (3) Finally, for the same reduced dataset, 3 repetitions of a 10-fold cross-validation process were applied to evaluate the selected variables through the performance of a Gradient Boosting classifier with the AUC metric. This three-step strategy is carried out through 10 iterations.

Trying to find a pattern, we ran the above process 4 times. Figure 6 shows the 14 selected features (with their importance hierarchy) that resulted in the highest mean AUC score of the 10 iterations, in 4 different runs. The red bars correspond to the features that were repeatedly chosen in every one of the 10 iterations for each run.

The students' perceived quality of the university teaching ("ACADEMIC_Q3"), followed by the GPA of the semester in which the survey was conducted ("GPA 202110") was the most relevant feature in all the runs conducted. The loss of attention during the synchronous activities ("SYNCH_ATT") caused by noise was one of the more relevant acoustic variables in all the runs.

Table 4 shows the Mean AUC score and the standard deviation of different models built using RFE with Gradient Boosting as the classifier. Although the performances are similar, the highest Mean AUC score corresponds to the 4th run.

Table 4. Mean AUC score and standard deviation (Sd) of the four algorithms used within the Recursive Feature Elimination method to obtain the relative importance of the variables. The AUC score is the area under the ROC (receiver operating characteristic) curve.

Run	Mean AUC Score	Sd
1	0.7827	0.0368
2	0.7823	0.0354
3	0.7819	0.0361
4	0.7828	0.0366

Comparing the 4 runs conducted, it is worth mentioning that the following features are present in all of them: "INT_PEOP_SYNCH", "SYNCH_ATT", "ACADEMIC_Q2", "ACADEMIC_Q3", "GPA_202010", "GPA_202020", "GPA_202110". However, for the sake of simplicity, the set of features with the highest Mean AUC score was used for the next steps, that is, the features of the 4th run.

The reduced dataset with the 14 selected features is now characterized by 9 columns representing the "perception of intensity and recurrence of interference", 2 columns representing "perception of teaching quality", and all 3 GPA rating columns. The 9 variables for the perception of intensity and the recurrence of interference were: the amount of interference of the domestic acoustic environment with autonomous-writing activities ("AUTO_WRI"); the recurrence of interference of domestic sounds with the loss of attention during synchronous activities ("SYNCH_ATT"); the amount of interference from the people in workspace with synchronous lessons ("INT_PEOP_SYNCH"); the recurrence of interference of the acoustic environment in keeping the thread of a class ("SYNCH_THR"); the amount of interference of the general domestic environment with an understanding of synchronous lessons ("INT_SYNCH"); the amount of interference of traffic noise with autonomous tasks ("AUTO_TRA"); the amount of interference of the animal noises with synchronous tasks ("SYNCH_ANI"); the amount of interference of voice-related noise with autonomous tasks ("AUTO_VOI"); the perceived loudness of TV sounds during academic activities ("LOUD_TV"). In total, 8 of these 9 features have to do with the inten-

sity and recurrence of interference from the inner environment of the student (domestic environment).

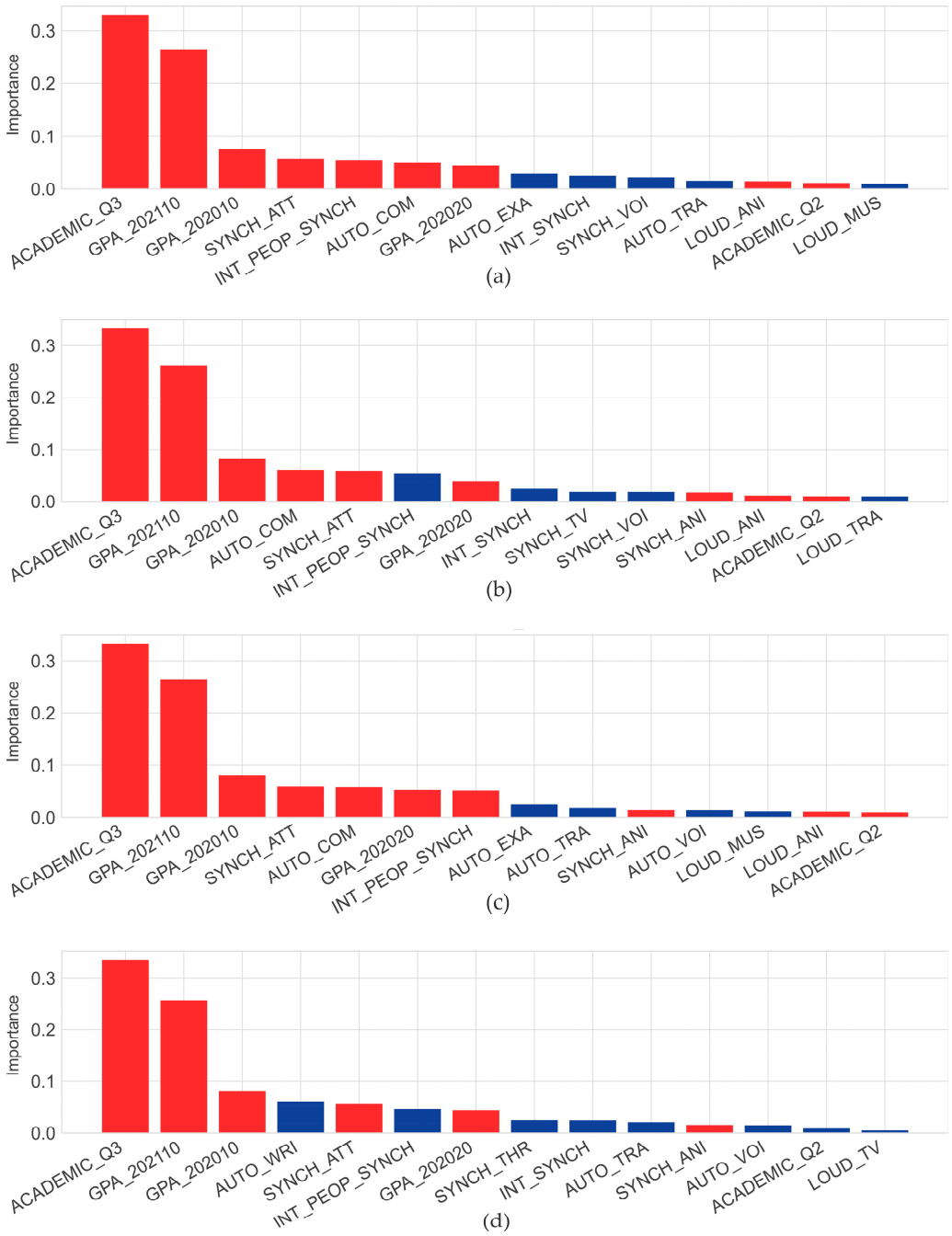

Figure 6. Relative importance of 14 variables after applying the Recursive Feature Elimination method to the whole dataset in 4 different runs. From top to bottom, run 1 to 4 (**a**–**d**).

It is noteworthy that four out of the five noise sources queried in the survey (traffic, animals, voices, TV/Radio/Household appliances) on different aspects (interference in autonomous and synchronous activities, loudness. . .) are included only once among the most relevant variables, so the model avoids duplicating the information of sound sources. The variables selected address interference in both autonomous (voices and traffic) and synchronous activities (animals), as well as the loudness of TV/Radio/Household appliances. A recent study conducted during the COVID lockdown considered that noises from

people at home, as well as traffic noise, are quite intrusive for working at home, which is in line with the results of our study [63].

The missing noise source is associated with music which appears not to be t relevant to that model, probably because many questions were asked about the interference of the sound sources, and students may not consider music to be an interference. This pertains to the outcomes of various studies examining the effects of listening to music as ambient sound. In recent research conducted by Krause et al., participants reported listening to music for emotional support and disengagement from problems [64]. Additionally, the same research group suggests that during the COVID-19 lockdown, listening to music was correlated with increased life satisfaction [9]. Music enhances the quality of work in professional settings and the task duration is extended when music is absent [65]. Furthermore, it was observed to boost efficiency and productivity during repetitive tasks. All this research led to the conclusion that (because normally the students themselves decide when and what kind of music to listen to) music can be seen as masking unwanted sound [63,66], and it is hardly considered to be a background disturbance.

There are also features regarding the extent of noise interference in activities. Regarding this, autonomous activities most notably affected include working in groups and writing, while synchronous activities are influenced by a diminished attention span and difficulties in following classes. This is in line with the study conducted by Torresin et al., which suggests that relaxation is important during reading activities to allow immersion and concentration without distractions in an ideal domestic acoustic environment [63].

3.5. Implementation of Classifiers

In this section, the reduced dataset, with the 14 selected features was used to calculate the performance of the models using three algorithms: Random Forest, Gradient Boosting, and SVM with Linear, Polynomial, and Radial Kernels.

Regarding the models, the calculation settings and how the convergence occurs will be defined in the following subsections. According to previous results, only the data used to calculate the SVM models were standardized.

3.5.1. Random Forest: Tuned Parameters

The graphs in Figure 7 represent the process of hyperparameter optimization. The AUC-ROC metric was used to evaluate each hyperparameter optimization. The mtyr hyperparameter refers to the number of variables that must be considered when searching for the best division in each node of the decision tree. Therefore, this exploration was based on testing various mtyr values (from 1 to 14). Figure 7a indicates which mtyr value obtained the highest metric (mtry = 2). Once this is completed, we proceeded to train the next hyperparameter (max nodes) with the same function and same criterion, but already using the optimized value of mtry. The maxnodes hyperparameter controls the maximum depth of the decision tree, which refers to the maximum number of levels (or nodes) from the root node to the leaves of the tree. Limiting the maximum depth of a tree can prevent overfitting the training data. Two intervals of different possible values for this hyperparameter were tested; a first range between 5 to 70, and a second one between 60 and 90. For both intervals, the best metric obtained was maxnode = 66. Therefore, to optimize the following hyperparameter, we use this value in the train() function.

The ntree hyperparameter specifies how many trees to build in the random forest. A larger number of trees generally improves the model's ability to generalize to new data and increases the model's robustness to overfitting. A range between 200 and 2000 was first tested. The value with the highest metric, in this case, was with ntrees = 1000.

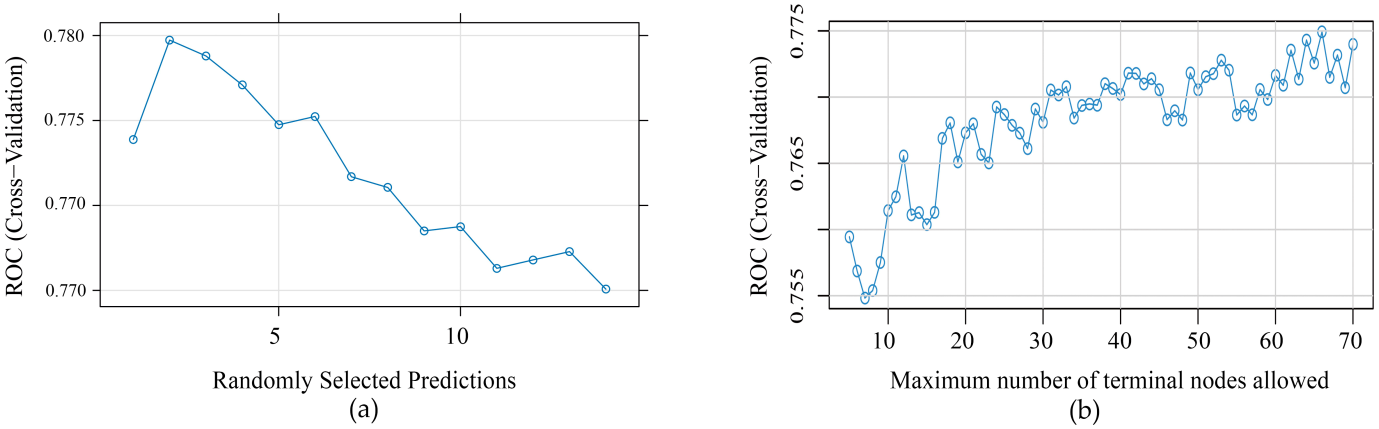

Figure 7. (**a**) Tuning of the mtry hyperparameter to obtain the optimal number of randomly selected predictions. (**b**) Maximum number of terminal nodes used to train the model.

3.5.2. Gradient Boosting: Tuned Parameters

Three hyperparameters were also tested. Due to the structure of the train() function, it is possible, in this case, to explore various values for the hyperparameters eta (learning rate) and n-round (number of iterations or training rounds that will be performed) during the model fitting process. The learning rate, also known as the step-down rate, controls the contribution of each tree to the overall model. A low learning rate means that each tree will have a smaller contribution to the final model and, therefore, require more trees in the assembly to obtain a more complex model. On the other hand, a high learning rate could result in overfitting. In each training round (n-round), the algorithm fits a new decision tree to the residual (difference between the current prediction and the actual value) of the existing model.

Figure 8a shows the tuning of the learning rate hyperparameter and Figure 8b shows the maximum depth used to fit the Gradient Boosting model. It can be appreciated that eta gives the same result for the tested values (as the whole line is represented in Figure 8a with the same green color), and considering that a high learning rate could result in overfitting, the lowest value was chosen for the optimization of the next hyperparameter (eta = 1×10^{-4}). But in the case of training rounds, the maximum value of AUC was given as 500. With these values, the Gradient Boosting model was optimized. Figure 8b shows that the maxdepth that obtained the best results was 9.

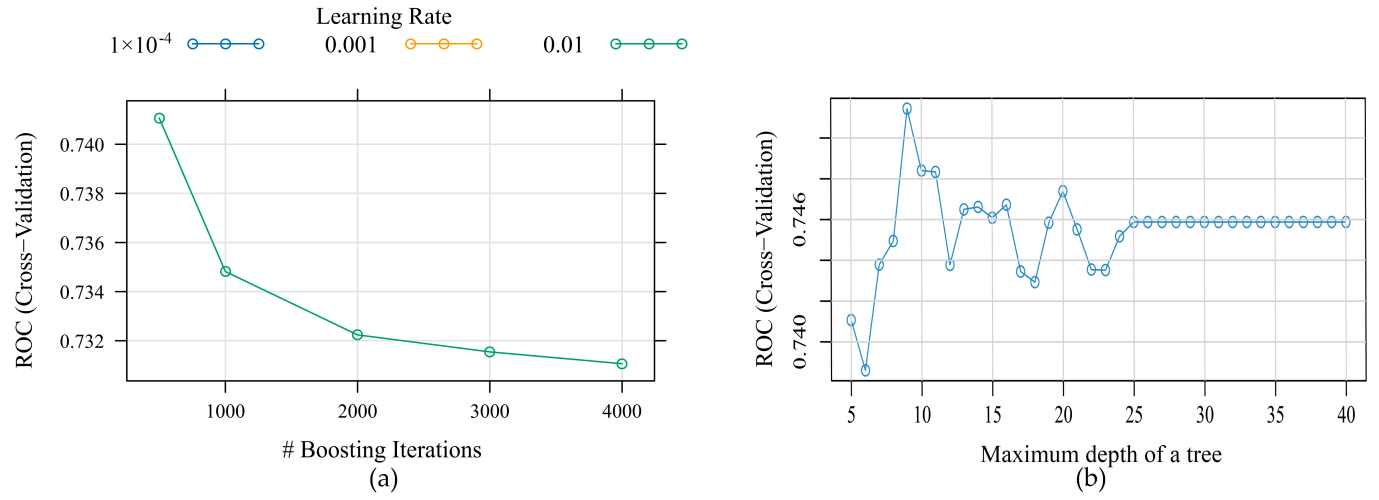

Figure 8. (**a**) Tuning of the learning rate hyperparameter, and (**b**) Maximum depth used to fit the Gradient Boosting model.

3.5.3. Support Vector Machine: Tuned Parameters

Three types of SVM are applied, each one with a different Kernel. The Linear Kernel does not transform the input data and it is suitable when the data is expected to be linearly separable in the original feature space. The parameter cost controls the balance between maximizing margin and minimizing classification error. Margin maximization finds a hyperplane that optimally separates different classes of data. The margin refers to the perpendicular distance between the decision hyperplane and the closest data points of the classes, which are called support vectors. For this model, the best cost was = 0.001. A cost value of 0.001 indicates that the SVM model penalizes classification errors relatively little compared to the importance it gives to margin maximization. The Polynomial Kernel transforms the input data into a higher dimensionality space using a polynomial expansion. The cost that leads to the highest AUC was 1 (Figure 9a). The degree indicates the degree of the polynomial that will be used to transform the data to a higher dimensionality feature space. For the Polynomial SVM model, the degree was set to 2. The Radial Kernel maps data into an infinite dimensional feature space. This Kernel is useful for separating non-linearly separable data and is the default Kernel in many cases. Sigma is a critical hyperparameter that controls the width of the Gaussian function used to calculate similarities between pairs of data points in the feature space. This parameter applies specifically to the Radial Kernel because of the way this Kernel computes similarities between pairs of data points in the feature space. The Sigma selected was 0.01, with cost = 1 (Figure 9b).

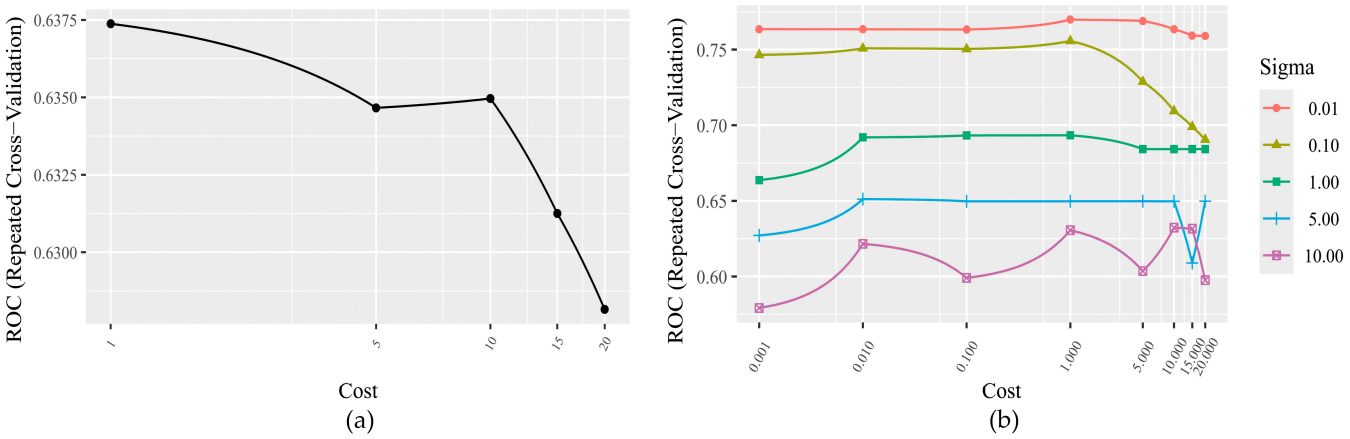

Figure 9. (**a**) Evaluation of the cost metric to calculate the support vector machine (SVM) with the Polynomial Kernel, (**b**). Evaluation of the sigma metric to calculate the SVM with the Radial Kernel.

Once the calculation settings have been outlined, the metrics used to evaluate the performance of the predictive models are described in the following section.

3.5.4. Models Metrics and Comparison

Table 5 shows the following metrics to evaluate the performance of the prediction model with 14 variables: Accuracy, McNemar's test *p*-value, Positive Predictive value, Negative Predictive value, and Mean AUC-Score.

The SVM with Radial Kernel is the model with better accuracy, showing a success rate of predictions of 78.7%, followed by the Random Forest model, with an accuracy = 77.94%. McNemar's test null hypothesis confirms that the performance of a baseline model built with random data and the prediction model evaluated are equal. McNemar's test *p*-values < 0.05 indicate that the null hypothesis can be rejected for all the models but the SVM with Polynomial Kernel. Therefore, Random Forest, Gradient Boosting, SVM with Linear kernel, and SVM with Radial Kernel models perform statistically significantly better than the baseline model; the SVM model with Polynomial Kernel does not perform better than a model with random predictions, so it will not be considered in the following explanations.

Table 5. Metrics of the performance of the models with 14 predictors. From left to right, Accuracy, Mcnemar's test *p*-value, Positive Predictive value, Negative Predictive value, and AUC score. The AUC score is the area under the ROC (receiver operating characteristic) curve.

Algorithm	Accuracy	McNemar *p*-Value:	Positive Pred Value	Negative Pred Value	AUC Score
Random Forest	0.7794	7.36×10^{-11}	0.7901	0.7143	0.7614
Gradient Boosting	0.7672	0.0015	0.8077	0.6145	0.7466
SVM (Linear)	0.7672	0.0051	0.8109	0.6111	0.7596
SVM (Polynomial)	0.7085	0.9335	0.8000	0.4822	0.7091
SVM (Radial)	0.7874	1.36×10^{-7}	0.8049	0.7023	0.7462

The Positive Predictive Value metric may be affected by the uneven distribution of classes. Given the high proportion of the positive class, the number of TP is high. Here, the role of the misclassified minority class (BAD class) FP is to diminish the ratio. For example, if we consider the metric for the Random Forest model, FP is relatively high (89 out of 424), but TP is much higher (335), due to the disproportion of positive cases (see Figure 10); thus, the ratio is still large. Nonetheless, this metric is basically about the positive class, and because the positive class is over-represented, the metric is slightly higher than the negative predictive ratio; however, the differences between Positive Predictive Values and Negative Predictive values are lower for the Random Forest model than for other algorithms, showing its good performance.

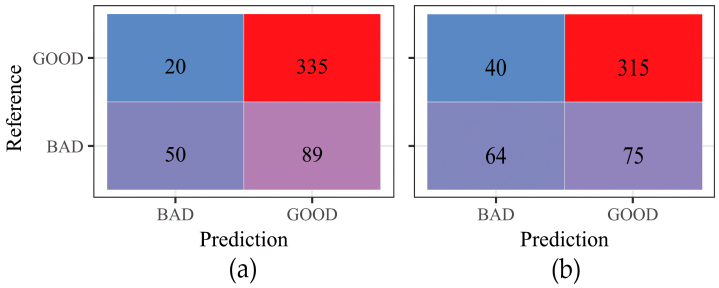

Figure 10. Confusion matrix of the following models with 14 predictors: (**a**) Random Forest, with 335 True positives (TP), 50 True negatives (TN), 89 False positives (FP), and 20 False negatives; (**b**) Gradient Boosting, with 306 TP, 64 TN, 75 FP, and 49 FN. The colors of each graph are a gradient between the minimum (light blue) and the maximum (red) number of predictions.

In this study, the uneven distribution of classes affects more other models than Random Forest regarding the comparison between Positive and Negative Predictive values (Table 5). For example, Gradient Boosting or SVM with Linear Kernel ones (0.8077 and 0.8109, respectively); consequently, high Positive Predictive values could be misleading these models. If we consider the SVM with Linear Kernel, the Positive Predictive value is more influenced by the majority class (GOOD class) and this can bias the metric towards a higher value. The number of TP is 313 out of 386 positive predictions (Figure 11), which leads to a high Positive Predictive value (0.8109). The Negative Predictive value (0.6111), however, is smaller, as the negative class is under-represented, remarking the existing difference between positive and negative predicted values.

Figures 12 and 13 show the AUC scores of the prediction models calculated with 14 predictors. The AUC score expresses the ability to accurately classify classes on a scale from 0.5 to 1, 0.5 being a random prediction, and 1 being a perfect discrimination ability. According to the criteria followed in this study, all the models show a good discrimination ability (AUC between 0.70 and 0.79) [61].

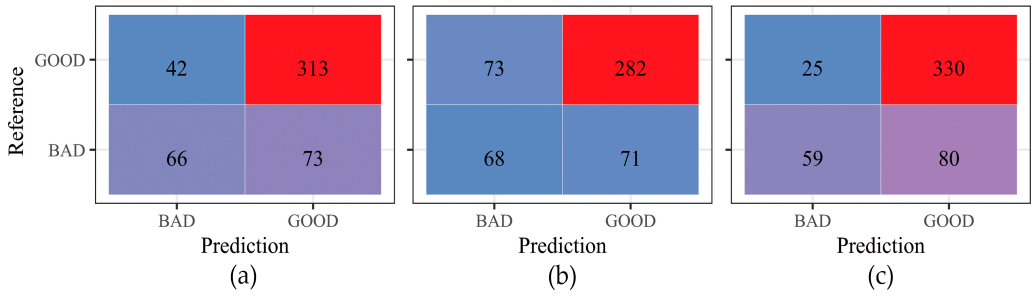

Figure 11. Confusion matrix of the following modes with 14 predictors: (**a**) Support Vector Machine (SVM) with Linear Kernel, with 313 True Positives (TP), 66 True Negatives (TN), 73 False positives (FP), and 42 False negatives; (**b**) SVM with Polynomial kernel, with 282 TP, 68 TN, 71 FP and 73 FN; (**c**) SVM with Radial Kernel, with 330 TP, 59 TN, 80 FP, and 25 FN. The colors of each graph are a gradient between the minimum (light blue) and the maximum (red) number of predictions.

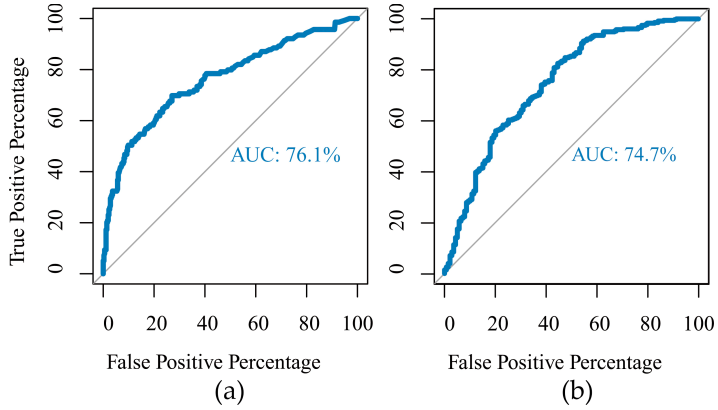

Figure 12. Area under the ROC (receiver operating characteristic) curve of the following models with 14 predictors: (**a**) Random Forest; (**b**) Gradient Boosting.

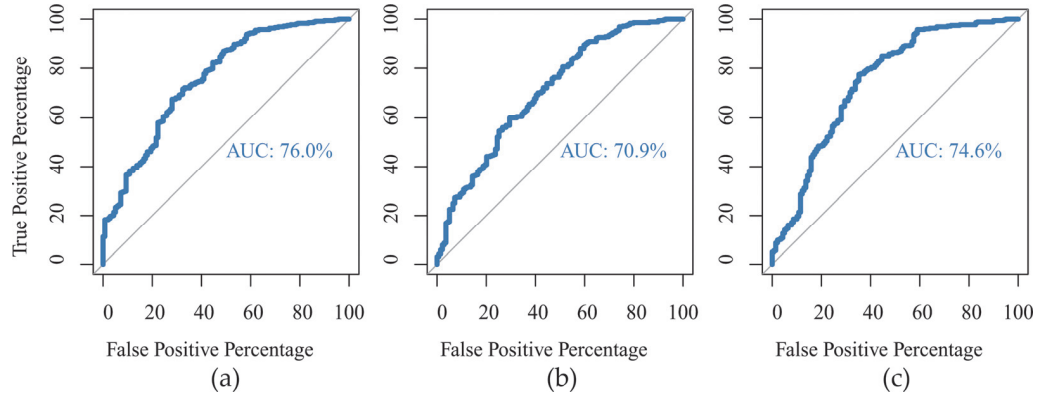

Figure 13. Area under the ROC (receiver operating characteristic) curve of the following models with 14 predictors: (**a**) Support Vector Machine (SVM) with Linear Kernel; (**b**) SVM with Polynomial Kernel; (**c**) SVM with Radial Kernel.

Remarkably, the differences between the models with 14 variables (Table 5) and the models with the optimal number of variables (Tables 2 and 3) in the Accuracy and Mean AUC Score metrics are minimal. It is also notable that, although the optimal number of variables that was obtained in the Recursive Feature Elimination phase was with the Gradient Boosting algorithm, this algorithm has not been the one that has obtained a better accuracy or Mean AUC score in the classifier implementation phase. This is because the results of the models may differ due to the stochastic nature of the algorithms and the evaluation process [40,67].

The comparison of the AUC score of the models calculated with the algorithms Random Forest, Gradient Boosting, and SVM using the optimal number of variables for each algorithm (34, 18 and 31, respectively with standardized data) and 14 variables leads to differences that range from −0.0031 (SVM with linear kernel) to 0.0071 (Random Forest). These are small differences, compared to the benefits of calculating machine learning-based models with the most relevant features: Firstly, by prioritizing the most relevant features, the model performance can be enhanced, thereby increasing the accuracy in predicting new, unseen data. Secondly, reducing the number of features can decrease the computational complexity of a model, resulting in quicker training and inference times. This is especially crucial for real-time or large-scale applications. Finally, focusing on essential features and eliminating noise or irrelevant data can enhance the interpretability of a model and of the predictor variables. This facilitates the understanding of the model's predictive mechanisms and identifying the key features driving those predictions [68–70].

A model with the global best-performed algorithm—Random Forest—was built, but using only the acoustic variables (9) out of the 14 used for the previous models, namely "AUTO_WRI", "SYNCH_ATT", "INT_PEOP_SYNCH", "SYNCH_THR", "INT_SYNCH", "AUTO_TRA", "SYNCH_ANI", "AUTO_VOI", and "LOUD_TV". This would allow us to know if the acoustic variables can be a factor to include when studying the perceived academic performance. The metrics of the calculated model are shown in Table 6. Figure 14 shows the confusion matrix and the AUC score of the model.

Table 6. Metrics of the performance of the model calculated only with 9 acoustic predictors. From left to right, Accuracy, McNemar's test *p*-value, Positive Predictive value, Negative Predictive value, and AUC score. The AUC score is the area under the ROC (receiver operating characteristic) curve.

Algorithm	Accuracy	McNemar *p*-Value:	Positive Pred Value	Negative Pred Value	AUC Score
Random Forest	0.7287	2.38×10^{-20}	0.7386	0.5806	0.6770

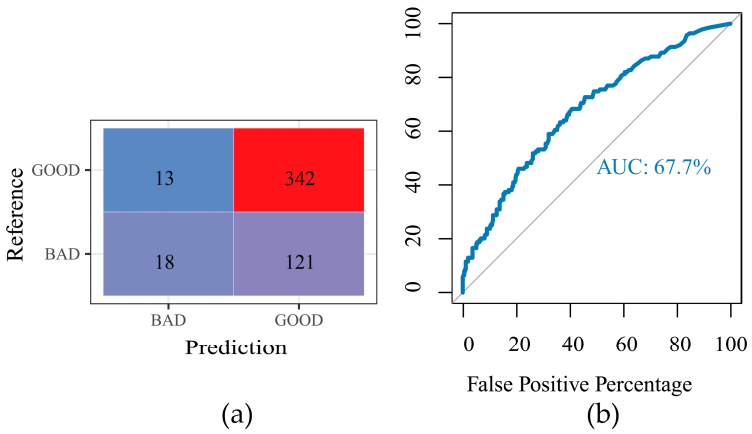

Figure 14. (**a**) Confusion matrix of the Random Forest calculated with 9 predictors, all of them acoustic variables, with 342 True Positives (TP), 18 True Negatives (TN), 121 False Positives (FP), and 13 False Negatives (FN); (**b**) Area under the ROC (receiver operating characteristic) curve of the Random Forest model calculated with 9 acoustic predictors. The colors of the confusion matrix graph are a gradient between the minimum (light blue) and the maximum (red) number of predictions.

McNemar *p*-value (2.38×10^{-20}) shows that the null hypothesis can be rejected for the model that only uses the perceived acoustic factors of the home environment as predictors, which leads to the conclusion that this model performs statistically significantly better than a model with random predictors. The AUC score is 0.68. Therefore, according to the chosen criterion [61], the model shows a moderate performance (AUC between 0.60 and 0.69). The model accuracy is 0.7287; consequently, the percentage of correct predictions is 72.87% only

with acoustic variables. This means that a model with 14 variables of different natures has only a 5% higher success rate than a model with only 9 acoustic variables; These are outstanding results considering the input variables, revealing that the acoustic environment plays an important role in perceived academic performance in online education modality.

Given the unique conditions that occurred during the COVID-19 lockdown, the home acoustic environment became somewhat chaotic. Household tasks (e.g., vacuuming, using the washing machine or dishwasher), and recreational activities (watching television, listening to music, playing video games) that produced high noise levels were carried out simultaneously with synchronous and autonomous academic tasks. As all family members had to coincide at home simultaneously, situations arose in which multiple family members had to carry out their work or academic tasks in the same space. Furthermore, flexible scheduling could lead to situations where some family members were studying or working while others were conducting leisure activities (for example, listening to music or cleaning their room). Since numerous studies demonstrate that noise interferes with intellectual tasks such as reading and listening comprehension, working memory, or mathematical solving problems [71–75], and also increases mental fatigue levels [76], it is understandable that students associate the acoustic home environment with their academic performance. Since domestic sounds may interfere with the development of academic activities and affect student's performance, creating a conducive learning environment with appropriate noise levels is crucial for academic success.

Comparing the performance of our prediction models with those of other studies presents challenges, since in our research, the target variable—self-reported academic performance (ACAD_PERFO)—is subjective, while in most of the others, the target variable is an objective measure like GPA or equivalent ratings [28,77–90]. We were interested in how students perceived their performance, which is totally different from actual academic achievements. Indeed, we expected a strong linear association between these two variables ($r > 0.8$), but our findings reveal that the Spearman correlation coefficient between self-perceived academic performance and the GPA is only $r = 0.172$ (Figure 1). The perception of academic performance (ACAD_PERFO) could vary from actual achievement due to its psychological nature influenced by individual expectations, family dynamics, and educational context, potentially involving cognitive biases [91]. Hence, comparing the performance metrics of models that predict two different target variables (subjective academic performance and GPA) may lead to misinterpretations of the results of our study, which we aim to avoid.

Teuber et al. [92] conducted a two-level regression model to predict the perceived academic performance, considering leisure activities, physical exercise, brief physical activity interruptions during study periods, and the characteristics of the student's home environment during the COVID pandemic lockdown. The perceived academic performance parameters were attention and study ability. In total, 57 students participated in the study. Marginal and conditional determination coefficients are the metrics to evaluate the outcomes for attention (0.126/0.514) and study ability (0.175/0.393). A study conducted by Azpiazu et al. [93] calculated a structural equation model to predict the perceived academic performance of secondary school students using a structural equation model. The predictor variables are integrated into the following groups: teacher support, peer support, resilience, positive affect, emotional engagement, and integration problems. Similarly, Oddermat et al. conducted a structural equation model to predict the perceived academic performance of adolescents and conducted descriptive statistics [94]. Jussila et al. conducted a study to predict the perceived academic performance of secondary school students calculating a logistic regression model [95] that used as predictors demographic, environmental, lifestyle, physical activity, active school transport, and leisure time factors. A total of 34,103 secondary school students participated in the study. They concluded that active school transport and leisure time were positively correlated with perceived academic performance. Similarly, Lee et al. conducted a multilinear regression analysis to predict the academic performance of university athletes, using R^2 to report the model

performance. 304 student-athletes from public and private universities in the United States participated in the study. The predictor variables were grit, life satisfaction, self-efficacy, anxiety, and NCAA divisions (top-tier athletes, and athletes with medium and limited economic resources), achieving a determination coefficient of $R^2 = 0.747$. Although the results and findings of the previous studies are promising in predicting perceived academic performance, given the disparities in academic modalities, predictor variables, and metrics utilized, comparisons of the outcomes with the ones of our study cannot be reported.

Ahmed and Rashidi predicted the perceived academic performance on face-to-face academic modality using simple and multivariate linear regression [96]. In total, 345 questionnaires were used to collect the information used in the study. The participants were university students from a metropolitan area. They calculated different predictive models with the following variables as unique factors (although some factors were perception scales composed of several variables): well-being (GB), Global Trait Emotional Intelligence Scale with 153 variables (GTEI) [97], Self-control (SC), Emotionality (EM), Sociability (SC), reduced Trait Emotional Intelligence Scale with 30 variables (TEIQue-SF) [98], and the general self-efficacy scale with 10 variables (GSE) [99]. The correlation coefficient (r) and the coefficient of determination (R^2) were the metrics used to explain the performance of the model. These are the correlation coefficients between the perceived academic performance and the previous variables: r(GB) = 0.250, r(GTEI) = 0.184, F(SC) = 0.288, r(EM) = 0.173, r(SC) = 0.096, r(TEIQue-SF) = 0.262, r(GSE) = 0.340. Some of the models with better correlation were composed of more than 10 variables. We have shown only the correlation coefficient r to have a baseline metric to compare the previous results with the ones of our study. The correlation coefficients between the perceived academic performance and the acoustic variables of our study with an absolute value $r > 0.25$, p-values < 0.001, and negative associations are: INT_PEOP_SYNCH (−0.253), SYNCH_ATT (−0.285), SYNCH_THR (−0.300), SYNCH_ANS (−0.273), AUTO_MAT (−0.256), AUTO_COM (−0.288), AUTO_WRI (−0.256), and AUTO_EXA (−0.291). The following acoustic variables present absolute correlation coefficients that ranges between 0.25 and 0.15 (p-value < 0.001): INT_PEOP_AUTO (−0.224), SYNCH_VOI (−0.197), SYNCH_TV (−0.167), SYNCH_ANI (−0.152), AUTO_VOI (−0.185), AUTO_TV (−0.165), SYNCH_NOT (−0.247), AUTO_WOR (−0.230), AUTO_GRA (−0.191), LOUD_VOI (−0.181), N_INSIDE (−0.155), INT_SYNCH (−0.220), and INT_AUTO (−0.201). Beyond the differences with the research of Ahmed and Rashidi, since distinct academic modalities and types of variables are being analyzed, the correlation coefficients of both studies are of the same order of magnitude, which demonstrates the importance of acoustic factors in the perception of the student's academic performance. This is even more notable considering that our sample size consists of 2477 cases compared to theirs, which has 345 cases. Typically, larger samples tend to result in lower, but more reliable, correlation coefficients.

Our research has shown how only 9 acoustic variables can predict, with a good success rate (73%), the perceived academic outcomes, which reflects the important role that the acoustic environment plays in the learning process. We do not intend to conclude from this that acoustic variables are the most important in predicting academic performance perception. However, the acoustic environment is indeed a factor that should be taken into account when assessing the online learning method from a holistic perspective. Another aspect to consider is that our study corresponds to the period of the COVID lockdown and that the results may evolve over time [100,101]. Technological advances are being implemented at a high speed, including new tools applied to education such as virtual reality or augmented reality [102,103]. From an acoustic point of view, the interference of this type of resource with the acoustic environment represents a challenge. To ensure the long-term sustainability of online education and guarantee equal opportunities in access to education, governments and educational institutions must prioritize the promotion of investment aid in infrastructure in homes that improves the insulation conditions of rooms, allows access to high-quality technology (such as noise-canceling headphones or other

appropriate acoustic technological tools, and internet resources), conduct noise awareness campaigns, and promote educational approaches that mitigate acoustic issues.

3.6. Research Limitations and Future Lines of Research

This study outlines certain limitations that deserve attention. Firstly, the utilization of ordinal variables, represented by numerical values for distinct categories, introduces a limitation. While these numerals denote the sequence of categories, they do not necessarily imply uniform intervals between them; subsequently, there is an assumption for consistent increments in the self-perception of academic performance levels that may lead to misinterpretations [104]. Secondly, our sample comprises university students exclusively from Universidad de Las Américas, and the study was conducted during the COVID-19 lockdown, both of which restrict the generalizability of our findings [100,101]. Thirdly, this study's aims to assess the impact of online learning modalities on self-rated academic performance inherently imposes limitations. Recognizing this objective as a constraint from the outset acknowledges that various other factors influencing student satisfaction remain unexplored. Finally, to simplify the distribution of responses and better understand the results, the dependent variable "ACAD_PERFO" was coded as "GOOD" (scores > 4) and "BAD" (scores $\leq$ 4). This results in information being lost since the dependent variable was coded in 7 classes. In conclusion, it is crucial to acknowledge these limitations to ensure a nuanced understanding of the study's outcomes.

Feature research may be conducted to understand how encoding the target variables affects the performance of the models. For example, the target variable can be coded into 7 classes, just as the original responses were coded, for a deeper understanding of the student's perception of their academic performance. However, this line of research has some weaknesses and should be thoroughly and cautiously evaluated, since considering more classes may generate a greater imbalance in the dataset. Another line of research would be to compare the impact of the acoustic environment on the perceived academic outcomes during (this has to be achieved using data compiled during the COVID-19 lockdown—not making a retrospective study—as acoustic memory is less robust than visual memory [105], and may introduce bias in the real acoustic perception) and after the COVID-19 lockdown, considering also hybrid and blended learning modalities.

4. Conclusions

The domestic environment and ICT are the fundamental media through which students communicate with both their instructors and peers in online education. Sounds, a fundamental part of the domestic environment, generate distractions and interferences that can play a pivotal role in this learning process. However, only a limited number of studies have been conducted to assess the impact of the acoustic environment on online teaching. This study examines how sound influences the self-perceived academic performance of university students during online education.

Different machine learning-based algorithms, in particular Random Forest, Gradient Boosting, and SVM were used to predict the perceived academic performance. The models were calculated using different predictor variables, most of them based on the acoustic environment, selected through their relative importance. Comparisons with the models calculated with the optimal number of variables for each algorithm and the reduced dataset were compared. The difference between the Accuracy score, which evaluates the percentage of correct predictions, is insignificant, which leads to the conclusion that a model composed of the most relevant variables gives an excellent performance, facilitates its interpretation, and reduces collinearities between variables. Based on the Accuracy score, the percentage of correct predictions varies between 76.72% (SVM with Linear Kernel and Gradient Boosting) and 78.74% (SVM with Radial Kernel). Based on the Mean AUC score, the percentage of correct predictions varies between 74.62% (SVM with Radial Kernel) and 76.14% (Random Forest), which involves a good discrimination ability of all the prediction models.

Relevant factors identified for student perception of their academic performance included noises interfering with reading, attention loss, and the disturbance caused by traffic, voices, animals, and TV/radio/household appliances.

The following answer the research questions of this study about the prediction of the perceived academic performance using acoustic variables:

- The model of fourteen variables, nine of them acoustic models, with better overall performance, is Random Forest with a success rate of 78%. While these values may seem modest, they are very good results, since most variables are related to the perceived acoustic environment. Furthermore, we reduced the forty initial variables to fourteen, which makes it easy to interpret the importance of each variable in the construction of the prediction model.
- It is noteworthy that four out of the five noise sources queried in the survey (traffic, animals, voices, and TV/Radio/Household appliances) on different aspects (interference in autonomous and synchronous activities, loudness. . .) are included only once among the most relevant variables; furthermore, the nine variables selected address interference in both autonomous (voices and traffic) and synchronous activities (animals), as well as the perceived loudness (of TV/Radio/Household appliances). Consequently, the acoustic variables that conform to the model, selected by the Recursive Feature Elimination method, avoid duplicating information.
- A model with only nine variables (all acoustics), out of the fourteen used for the previous model, obtained a success rate of 73% only 5% lower than a model that included, among others, variables as important as the real academic performance GPA or the perceived academic quality. Therefore, these outcomes show the important role of the acoustic variables in the perception of the student's academic performance.

Addressing noise-related issues, such as unexpected acoustic events and background noise, can potentially enhance the effectiveness of online learning and increase students' self-perception of their performance. To ensure the long-term sustainability of online education and ensure equitable access to education, governments and educational institutions need to prioritize investment in home infrastructure. This includes improving room insulation conditions, providing access to high-quality technology such as noise-canceling headphones and appropriate technological tools, as well as ensuring a high-quality internet connection. Additionally, conducting noise awareness campaigns and implementing educational approaches that address acoustic issues are other essential steps in the process of stabilization of and the adaptation to the new online learning modalities.

Author Contributions: Conceptualization, V.P.-R.; methodology, V.P.-R., A.M.D.-M., C.M.L.-Á. and G.C.; software, V.P.-R., C.M.L.-Á. and G.C.; validation, V.P.-R., C.M.L.-Á. and G.C.; formal analysis, V.P.-R., C.M.L.-Á. and G.C.; investigation, V.P.-R., A.M.D.-M., C.M.L.-Á. and G.C.; resources, V.P.-R., C.M.L.-Á. and G.C.; data curation, V.P.-R.; writing—original draft preparation, V.P.-R., C.M.L.-Á. and G.C.; writing—review and editing, V.P.-R., C.M.L.-Á., A.M.D.-M., G.C. and R.H.-M.; visualization, V.P.-R., G.C. and R.H.-M.; supervision, V.P.-R., C.M.L.-Á., A.M.D.-M., G.C. and R.H.-M.; project administration, V.P.-R.; funding acquisition, V.P.-R. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Universidad de Las Américas (UDLA), and was developed under the research project with the reference code SOA.VPR.20.03 (VII Call for research projects of the UDLA).

Institutional Review Board Statement: The Committee of Ethics in Human Research at the Universidad de Las Americas (CEISH-UDLA) determined that ethical approval was not deemed necessary for this study. This decision was based on the absence of any potential risks to the participants and their personal data, as the information provided was anonymized to protect confidentiality.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The data used for the present study will be made available on a reasonable request by contacting the corresponding author.

Acknowledgments: The authors would like to express our sincere gratitude for the selfless collaboration of all the students who participated in the study. We also wish to express our appreciation for the valuable contributions to the questionnaire's development provided by Paula Hidalgo Andrade, Carlos Alberto Hermosa, and Clara Patricia Paz from the Department of Psychology at UDLA. Additionally, our gratitude extends to Luis Antonio Vaca Hinojosa, from the Information Intelligence Department, for facilitating the data compilation. Special thanks are also due to Marlena León Mendoza, from the Academic Vice-rectorate, and Tannya Lorena Lozada for granting the necessary authorizations to promote the survey through the official networks of UDLA. Lastly, we appreciate Lien Gonzalez, from the Research Department Directorate, for her invaluable support in the project development.

Conflicts of Interest: The authors declare no conflicts of interest.

Appendix A

Table A1. Inquiries regarding the acoustic surroundings of students within their residences applied in the current research, along with their corresponding abbreviations.

Questions	Measurement Scale
1. In the last month, how do you think your academic performance has been? (ACAD_PERFO)	Very bad (1), Excellent (7)
2. In the last month, how loud do you think the following types of sounds have been during your academic activities (in general)? • Animals (LOUD_ANI) • Music (LOUD_MUS) • Traffic (LOUD_TRA) • TV/Radio/Household appliances (LOUD_TV) • Voices (LOUD_VOI)	I did not hear them (0), Very low (1), Neither high nor low (3), Very high (5)
3. In the last month, how much have the following types of sounds interfered with synchronous lessons? • Animals (SYNCH_ANI) • Music (SYNCH_MUS) • Traffic (SYNCH_TRA) • TV/Radio/Household appliances (SYNCH_TV) • Voices (SYNCH_VOI)	I did not hear them (0), They did not interfere (1), Extremely (5)
4. In the last month, how much have the following types of sounds interfered with your autonomous tasks? • Animals (AUTO_ANI) • Music (AUTO_MUS) • Traffic (AUTO_TRA) • TV/Radio/Household appliances (AUTO_TV) • Voices (AUTO_VOI)	I did not hear them (0), They did not interfere (1), Extremely (5)
5. In this last month, how often did the sounds of your home environment during synchronous lessons…? • Made you lose your attention? (SYNCH_ATT) • Caused you to not hear clearly? (SYNCH_NOT) • Made you lose the thread of the class? (SYNCH_THR) • Prevented you from answering the questions in time? (SYNCH_ANS)	Never (1), Very often (5)

Table A1. *Cont.*

Questions	Measurement Scale
6. In the last month, how much have the sounds of your domestic environment interfered with the following autonomous tasks: . . .? • Problem-solving (AUTO_MAT) • Working in groups (AUTO_WOR) • Comprehensive reading (AUTO_COM) • Writing (AUTO_WRI) • Making graphs/diagrams/drawings/models (AUTO_GRA) • Exam (practical-theoretical) (AUTO_EXA)	They did not interfere (1), Extremely (5)
7. In the last month, how much have the other people you share a workspace with interfered while attending synchronous lessons? (INT_PEOP_SYNCH)	They did not interfere (1), Extremely (5)
8. In the last month, how much have the other people you share a workspace with interfered while performing your autonomous tasks? (INT_PEOP_AUTO)	They did not interfere (1), Extremely (5)
9. In the last month, to what extent has the noise bothered you in your academic activities. . . • Coming from inside your home? (N_INSIDE) • Coming from outside your home? (N_OUTSIDE)	They did not interfere (1), Extremely (5)
10. In the last month, how much have the sounds of your domestic environment interfered with. . .? • Your synchronous lessons? (INT_SYNCH) • Your autonomous tasks? (INT_AUTO)	They did not interfere (1), Extremely (5)
11. In the last month, how often have you experienced audio connection problems attributable to your internet provider? (INTER_AUDIO)	I did not have problems (0), They did not disturb me (1), Extremely (5)
12. How long would you be able to endure a noisy environment? (ENDURE_NOI)	Just a little (1), A lot (5)
13. Order the following factors according to the degree of interference in your academic performance in the last month, the first being the one that has interfered the most. • Comfort of your home environment (ORDER_IN_1, not used in this study) • Acoustic environment (ORDER_IN_2) • Thermal environment (ORDER_IN_3, not used in this study) • Internet service quality (ORDER_IN_4, not used in this study) • Stress (ORDER_IN_5, not used in this study)	
14. How do you consider the quality of teaching at UDLA • before the pandemic, semester 202010 (ACADEMIC_Q1) • the first pandemic semester 202020 (ACADEMIC_Q2) • the second pandemic semester 202110 (ACADEMIC_Q3) *	Very bad (1), Neither bad nor good (3), Very good (5)

* Semester in which the survey was conducted.

Table A2. GPA of the semester in which the survey was conducted GPA 202110 and two previous semesters.

Academic Data—Marks	Measurement Scale
1. Average marks • before the pandemic, semester 202010 (GPA 202010) • the first pandemic semester 202020 (GPA 202020) • the second pandemic semester 202110 (GPA 202110) *	1 to 10

* Semester in which the survey was conducted.

References

1. Palacios Hidalgo, F.J.; Huertas Abril, C.A.; Gómez Parra, M.E. MOOCs: Origins, Concept and Didactic Applications: A Systematic Review of the Literature (2012–2019). *Technol. Knowl. Learn.* **2020**, *25*, 853–879. [CrossRef]
2. Metz, K. Benefits of Online Courses in Career and Technical Education. *Tech. Connect. Educ. Careers* **2010**, *85*, 20–23.
3. Bozkurt, A.; Zawacki-Richter, O. Trends and Patterns in Massive Open Online Courses: Review and Content Analysis of Research on MOOCs. *Int. Rev. Res. Open Distrib. Learn.* **2017**, *18*, 119–147.
4. Stansfield, M.; Mclellan, E.; Connolly, T. Enhancing Student Performance in Online Learning and Traditional Face-to-Face Class Delivery. *J. Inf. Technol. Educ. Res.* **2004**, *3*, 173–188. [CrossRef] [PubMed]
5. Daymont, T.; Blau, G. Student Performance in Online and Traditional Sections of an Undergraduate Management Course. *J. Behav. Appl. Manag.* **2008**, *9*, 275–294. [CrossRef]
6. Estelami, H. Determining the Drivers of Student Performance in Online Business Courses. *Am. J. Bus. Educ.* **2013**, *7*, 79–92. [CrossRef]
7. Bir, D.D. Comparison of Academic Performance of Students in Online vs. Traditional Engineering Course. *Eur. J. Open, Distance E-Learn.* **2019**, *22*, 1–13. [CrossRef]
8. Rebucas-Estacio, R.; CAllanta-Raga, R. Analyzing Students Online Learning Behavior in Blended Courses Using Moodle. *Asian Assoc. Open Univ. J.* **2017**, *12*, 52–68. [CrossRef]
9. Krause, A.E.; Dimmock, J.; Rebar, A.L.; Jackson, B. Music Listening Predicted Improved Life Satisfaction in University Students During Early Stages of the COVID-19 Pandemic. *Front. Psychol.* **2021**, *11*, 631033. [CrossRef] [PubMed]
10. Gopal, R.; Singh, V.; Aggarwal, A. Impact of Online Classes on the Satisfaction and Performance of Students during the Pandemic Period. *Educ. Inf. Technol.* **2021**, *26*, 6923–6947. [CrossRef]
11. Puyana-Romero, V.; Díaz-Marquez, Á.M.; Ciaburro, G.; Hernandez-Molina, R. The Acoustic Environment and University Students' Satisfaction with the Online Education Method during the COVID-19 Lockdown. *Int. J. Environ. Res. Public Health* **2023**, *20*, 709. [CrossRef] [PubMed]
12. Logan, E.; Augustyniak, R.; Rees, A. Distance Education as Different Education: A Student-Centered Investigation of Distance Learning Experience. *J. Educ. Libr. Inf. Sci.* **2002**, *43*, 32–42. [CrossRef]
13. Papadakis, S. MOOCs 2012–2022: An Overview Methods Study Design. *Adv. Mob. Learn. Educ. Res.* **2023**, *3*, 682–693. [CrossRef]
14. Damián-Chávez, M.M.; Ledesma-Coronado, P.E.; Drexel-Romo, M.; Ibarra-Zárate, D.I.; Alonso-Valerdi, L.M. Physiology & Behavior Environmental Noise at Library Learning Commons Affects Student Performance and Electrophysiological Functioning. *Physiol. Behav.* **2021**, *241*, 113563. [CrossRef] [PubMed]
15. Buchari, B.; Matondang, N. The Impact of Noise Level on Students' Learning Performance at State Elementary School in Medan. *AIP Conf. Proc.* **2017**, *1855*, 040002. [CrossRef]
16. Nelson, P.B.; Soli, S. Acoustical Barriers to Learning: Children at Risk in Every Classroom. *Lang. Speech Hear. Serv. Sch.* **2000**, *31*, 356–361. [CrossRef] [PubMed]
17. Choi, Y. The Intelligibility of Speech in University Classrooms during Lectures. *Appl. Acoust.* **2020**, *162*, 107211. [CrossRef]
18. Shield, B.M.; Dockrell, J.E. The Effects of Environmental and Classroom Noise on the Academic Attainments of Primary School Children. *J. Acoust. Soc. Am.* **2008**, *123*, 133–144. [CrossRef] [PubMed]
19. World Health Organization; European Union. *Guidelines for Community Noise. Document References MNB-1Q DOC2*; World Health Organization: Geneva, Switzerland, 1999; Volume 5(2)(a).
20. Carroll, N. European Journal of Higher Education E-Learning—The McDonaldization of Education. *Eur. J. High. Educ.* **2014**, *3*, 342–356. [CrossRef]
21. Wladis, C.; Conway, K.M.; Hachey, A.C. The Online STEM Classroom—Who Succeeds? An Exploration of the Impact of Ethnicity, Gender, and Non-Traditional Student Characteristics in the Community College Context. *Community Coll. Rev.* **2015**, *43*, 142–164. [CrossRef]
22. Clark-Ibáñez, M.; Scott, L. Learning to Teach Online. *Teach. Sociol.* **2008**, *36*, 34–41. [CrossRef]
23. Driscoll, A.; Jicha, K.; Hunt, A.N.; Tichavsky, L.; Thompson, G. Can Online Courses Deliver In-Class Results? A Comparison of Student Performance and Satisfaction in an Online versus a Face-to-Face Introductory Sociology Course. *Teach. Sociol.* **2012**, *40*, 312–331. [CrossRef]

24. Paul, J.; Jefferson, F. A Comparative Analysis of Student Performance in an Online vs. Face-to-Face Environmental Science Course From 2009 to 2016. *Front. Comput. Sci.* **2019**, *1*, 472525. [CrossRef]
25. González-Gómez, D.; Jeong, J.S.; Airado Rodríguez, D.; Cañada-Cañada, F. Performance and Perception in the Flipped Learning Model: An Initial Approach to Evaluate the Effectiveness of a New Teaching Methodology in a General Science Classroom. *J. Sci. Educ. Technol.* **2016**, *25*, 450–459. [CrossRef]
26. Pei, L.; Wu, H. Does Online Learning Work Better than Offline Learning in Undergraduate Medical Education? A Systematic Review and Meta-Analysis. *Med. Educ. Online* **2019**, *24*, 1666538. [CrossRef] [PubMed]
27. Lockman, A.S.; Schirmer, B.R. Online Instruction in Higher Education: Promising, Research-Based, and Evidence Based Practices 3. Themes in the Research Literature on Online Learning. *J. Educ. e-Learn. Res.* **2020**, *7*, 130–152. [CrossRef]
28. Qiu, F.; Zhang, G.; Sheng, X.; Jiang, L.; Zhu, L.; Xiang, Q.; Jiang, B.; Chen, P.-k. Predicting Students' Performance in e-Learning Using Learning Process and Behaviour Data. *Sci. Rep.* **2022**, *12*, 453. [CrossRef]
29. Aydoğdu, Ş. Predicting Student Final Performance Using Artificial Neural Networks in Online Learning Environments. *Educ. Inf. Technol.* **2020**, *25*, 1913–1927. [CrossRef]
30. Alshabandar, R.; Hussain, A.; Keight, R.; Khan, W. Students Performance Prediction in Online Courses Using Machine Learning Algorithms. In Proceedings of the 2020 International Joint Conference on Neural Networks, Glasgow, UK, 19–24 July 2020. [CrossRef]
31. Segura, M.; Mello, J. Machine Learning Prediction of University Student Dropout: Does Preference Play a Key Role? *Mathematics* **2022**, *10*, 3359. [CrossRef]
32. Regnier, J.; Shafer, E.; Sobiesk, E.; Stave, N.; Haynes, M. From Crisis to Opportunity: Practices and Technologies for a More Effective Post-COVID Classroom. *Educ. Inf. Technol.* **2023**, *29*, 5981–6003. [CrossRef]
33. Bashir, A.; Bashir, S.; Rana, K.; Lambert, P.; Vernallis, A. Post-COVID-19 Adaptations; the Shifts Towards Online Learning, Hybrid Course Delivery and the Implications for Biosciences Courses in the Higher Education Setting. *Front. Educ.* **2021**, *6*, 711619. [CrossRef]
34. Arday, J. COVID-19 and Higher Education: The Times They Are A'Changin. *Educ. Rev.* **2022**, *74*, 365–377. [CrossRef]
35. Akoglu, H. User's Guide to Correlation Coefficients. *Turk. J. Emerg. Med.* **2018**, *18*, 91–93. [CrossRef] [PubMed]
36. Miot, H.A. Analysis of Ordinal Data in Clinical and Experimental Studies. *J. Vasc. Bras.* **2020**, *19*, e20200185. [CrossRef] [PubMed]
37. Kulkarni, A.; Chong, D.; Batarseh, F.A. 5—Foundations of Data Imbalance and Solutions for a Data Democracy. In *Data Democracy*; Batarseh, F.A., Yang, R., Eds.; Academic Press: Cambridge, MA, USA, 2020; pp. 83–106. [CrossRef]
38. He, H.; Ma, Y. *Imbalanced Learning: Foundations, Algorithms, and Applications*, 1st ed.; Wiley-IEEE Press: Hoboken, NJ, USA, 2013.
39. Google for Developer. Imbalance Data. Available online: https://developers.google.com/machine-learning/data-prep/construct/sampling-splitting/imbalanced-data?hl=es-419 (accessed on 7 April 2023).
40. Brownlee, J. (Ed.) *Machine Learning Mastering with R, v. 1.12.*; Guiding Tech Media: Melbourne, Australia, 2016.
41. Baranwal, A.; Bagwe, B.R.; Vanitha, M. Machine Learning in Python. *J. Mach. Learn. Res.* **2019**, *12*, 128–154. [CrossRef]
42. Lemaître, G.; Nogueira, F.; Aridas, C.K. Imbalanced-Learn: A Python Toolbox to Tackle the Curse of Imbalanced Datasets in Machine Learning. *J. Mach. Learn. Res.* **2017**, *18*, 1–5.
43. Kuhn, M.; Johnson, K. Measuring Performance in Classification Models. In *Applied Predictive Modeling*; Kuhn, M., Johnson, K., Eds.; Springer: New York, NY, USA, 2013; pp. 247–273. [CrossRef]
44. Liu, Y.; Li, Y.; Xie, D. Implications of Imbalanced Datasets for Empirical ROC-AUC Estimation in Binary Classification Tasks. *J. Stat. Comput. Simul.* **2024**, *94*, 183–203. [CrossRef]
45. Priyatno, A.M.; Widiyaningtyas, T. A Systematic Literature Review: Recursive Feature Elimination Algorithms. *JITK (Jurnal Ilmu Pengetah. Teknol. Komputer)* **2024**, *9*, 196–207. [CrossRef]
46. Kuhn, M.; Johnson, K. An Introduction to Feature Selection. In *Applied Predictive Modeling*; Kuhn, M., Johnson, K., Eds.; Springer: New York, NY, USA, 2013; pp. 487–519. [CrossRef]
47. Butcher, B.; Smith, B.J. Feature Engineering and Selection: A Practical Approach for Predictive Models. *Am. Stat.* **2020**, *74*, 308–309. [CrossRef]
48. Raschka, S. *Machine Learning Q and AI*; No Starch Press: San Francisco, CA, USA, 2024.
49. Luor, D.-C. A Comparative Assessment of Data Standardization on Support Vector Machine for Classification Problem. *Intell. Data Anal.* **2015**, *19*, 529–546. [CrossRef]
50. Ganganwar, V. An Overview of Classification Algorithms for Imbalanced Datasets. *Int. J. Emerg. Technol. Adv. Eng.* **2012**, *2*, 42–47.
51. Breiman, L.; Friedman, J.; Olshen, R.A.; Stone, C.J. *Classification and Regression Trees*, 1st ed.; Chapman and Hall/CRC: New York, NY, USA, 1984; Volume 5. [CrossRef]
52. Rainio, O. Evaluation Metrics and Statistical Tests for Machine Learning. *Sci. Rep.* **2024**, *14*, 6086. [CrossRef] [PubMed]
53. García-Balboa, J.L.; Alba-Fernández, M.V.; Ariza-López, F.J.; Rodríguez-Avi, J. Analysis of Thematic Similarity Using Confusion Matrices. *ISPRS Int. J. Geo-Inf.* **2018**, *7*, 233. [CrossRef]
54. Das, C.; Sahoo, A.K.; Pradhan, C. Chapter 12—Multicriteria Recommender System Using Different Approaches. In *Cognitive Data Science in Sustainable Computing*; Mishra, S., Tripathy, H.K., Mallick, P.K., Sangaiah, A.K., Chae, G.-S., Eds.; Academic Press: Cambridge, MA, USA, 2022; pp. 259–277. [CrossRef]
55. Pembury Smith, M.Q.R.; Ruxton, G.D. Effective Use of the McNemar Test. *Behav. Ecol. Sociobiol.* **2020**, *74*, 133. [CrossRef]

56. Safari, S.; Baratloo, A.; Elfil, M.; Negida, A. Evidence Based Emergency Medicine Part 2: Positive and Negative Predictive Values of Diagnostic Tests. *Emergency* **2015**, *3*, 87–88.
57. Fawcett, T. ROC Graphs: Notes and Practical Considerations for Researchers. *Pattern Recognit. Lett.* **2004**, *31*, 1–38.
58. Zweig, M.H.; Campbell, G. Receiver-Operating Characteristic (ROC) Plots: A Fundamental Evaluation Tool in Clinical Medicine. *Clin. Chem.* **1993**, *39*, 561–577. [CrossRef] [PubMed]
59. Jones, C.M.; Athanasiou, T. Summary Receiver Operating Characteristic Curve Analysis Techniques in the Evaluation of Diagnostic Tests. *Ann. Thorac. Surg.* **2005**, *79*, 16–20. [CrossRef] [PubMed]
60. de Hond, A.A.H.; Steyerberg, E.W.; van Calster, B. Interpreting Area under the Receiver Operating Characteristic Curve. *Lancet Digit. Health* **2022**, *4*, e853–e855. [CrossRef]
61. Nassar, A.P.; Mocelin, A.O.; Nunes, A.L.B.; Giannini, F.P.; Brauer, L.; Andrade, F.M.; Dias, C.A. Caution When Using Prognostic Models: A Prospective Comparison of 3 Recent Prognostic Models. *J. Crit. Care* **2012**, *27*, e1–e423. [CrossRef]
62. Kuckartz, U.; Rädiker, S.; Ebert, T.; Schehl, J. *Statistik, Eine Verständliche Einführung*; Springer: Berlin/Heidelberg, Germany, 2013. [CrossRef]
63. Torresin, S.; Ratcliffe, E.; Aletta, F.; Albatici, R.; Babich, F.; Oberman, T.; Kang, J. The Actual and Ideal Indoor Soundscape for Work, Relaxation, Physical and Sexual Activity at Home: A Case Study during the COVID-19 Lockdown in London. *Front. Psychol.* **2022**, *13*, 1038303. [CrossRef]
64. Krause, A.E.; Scott, W.G.; Flynn, S.; Foong, B.; Goh, K.; Wake, S.; Miller, D.; Garvey, D. Listening to Music to Cope with Everyday Stressors. *Music. Sci.* **2021**, *27*, 176–192. [CrossRef]
65. Lesiuk, T. The Effect of Music Listening on Work Performance. *Psychol. Music* **2005**, *33*, 173–191. [CrossRef]
66. Nilsson, M.E.; Alvarsson, J.; Rådsten-Ekman, M.; Bolin, K. Auditory Masking of Wanted and Unwanted Sounds in a City Park. *Noise Control Eng. J.* **2010**, *58*, 524–531. [CrossRef]
67. Ciaburro, G. *MATLAB for Machine Learning*; Packt Publishing: Birmingham, UK, 2017.
68. Pudjihartono, N.; Fadason, T.; Kempa-Liehr, A.W.; O'Sullivan, J.M. A Review of Feature Selection Methods for Machine Learning-Based Disease Risk Prediction. *Front. Bioinform.* **2022**, *2*, 927312. [CrossRef] [PubMed]
69. Noroozi, Z.; Orooji, A.; Erfannia, L. Analyzing the Impact of Feature Selection Methods on Machine Learning Algorithms for Heart Disease Prediction. *Sci. Rep.* **2023**, *13*, 22588. [CrossRef]
70. Puyana-Romero, V.; Maffei, L.; Brambilla, G.; Ciaburro, G. Modelling the Soundscape Quality of Urban Waterfronts by Artificial Neural Networks. *Appl. Acoust.* **2016**, *111*, 121–128. [CrossRef]
71. Hétu, R.; Truchon-Gagnon, C.; Bilodeau, S.A. Problems of Noise in School Settings: A Review of Literature and the Results of an Exploratory Study. *J. Speech-Lang. Pathol. Audiol.* **1990**, *14*, 31–39.
72. Shield, B.; Dockrell, J.E. External and Internal Noise Surveys of London Primary Schools. *J. Acoust. Soc. Am.* **2004**, *115*, 730–738. [CrossRef]
73. Caviola, S.; Visentin, C.; Borella, E.; Mammarella, I.; Prodi, N. Out of the Noise: Effects of Sound Environment on Maths Performance in Middle-School Students. *J. Environ. Psychol.* **2021**, *73*, 101552. [CrossRef]
74. Nagaraj, N.K. Effect of Auditory Distraction on Working Memory, Attention Switching, and Listening Comprehension. *Audiol. Res.* **2021**, *11*, 227–243. [CrossRef] [PubMed]
75. Liu, C.; Zang, Q.; Li, J.; Pan, X.; Dai, H.; Gao, W. The Effect of the Acoustic Environment of Learning Spaces on Students' Learning Efficiency: A Review. *J. Build. Eng.* **2023**, *79*, 107911. [CrossRef]
76. Doctora, A.L.S.; Perez, W.D.D.; Vasquez, A.B.; Gumasing, M.J.J. Relationship of Noise Level to the Mental Fatigue Level of Students: A Case Study during Online Classes. In Proceedings of the International Conference on Industrial Engineering and Operations Management, Rome, Italy, 2–5 August 2021; pp. 1378–1386.
77. Khan, A.; Ghosh, S.K. Student Performance Analysis and Prediction in Classroom Learning: A Review of Educational Data Mining Studies. In *Education and Information Technologies*; Springer: Berlin/Heidelberg, Germany, 2021; Volume 26. [CrossRef]
78. Shou, Z.; Xie, M.; Mo, J.; Zhang, H. Predicting Student Performance in Online Learning: A Multidimensional Time-Series Data Analysis Approach. *Appl. Sci.* **2024**, *14*, 2522. [CrossRef]
79. Ismail, N.H.; Ahmad, F.; Aziz, A.A. Implementing WEKA as a Data Mining Tool to Analyze Students' Academic Performances Using Naïve Bayes Classifier. In Proceedings of the UniSZA Postgraduate Research Conference, Kuala Terengganu, Malaysia, 7–8 September 2013; pp. 855–863. [CrossRef]
80. Pandey, M.; Kumar Sharma, V. A Decision Tree Algorithm Pertaining to the Student Performance Analysis and Prediction. *Int. J. Comput. Appl.* **2013**, *61*, 1–5. [CrossRef]
81. Yang, S.J.H.; Lu, O.H.T.; Huang, A.Y.Q.; Huang, J.C.H.; Ogata, H.; Lin, A.J.Q. Predicting Students' Academic Performance Using Multiple Linear Regression and Principal Component Analysis. *J. Inf. Process.* **2018**, *26*, 170–176. [CrossRef]
82. Nedeva, V.; Pehlivanova, T. Students' Performance Analyses Using Machine Learning Algorithms in WEKA. *IOP Conf. Ser. Mater. Sci. Eng.* **2021**, *1031*, 012061. [CrossRef]
83. Al-Barrak, M.A.; Al-Razgan, M. Predicting Students Final GPA Using Decision Trees: A Case Study. *Int. J. Inf. Educ. Technol.* **2016**, *6*, 528–533. [CrossRef]
84. Folorunso, S.O.; Farhaoui, Y.; Adigun, I.P.; Imoize, A.L.; Awotunde, J.B. Prediction of Student's Academic Performance Using Learning Analytics. In *Artificial Intelligence, Data Science and Applications*; Farhaoui, Y., Hussain, A., Saba, T., Taherdoost, H., Verma, A., Eds.; Springer Nature: Cham, Switzerland, 2024; pp. 314–325.

85. Farooq, U.; Naseem, S.; Mahmood, T.; Li, J.; Rehman, A.; Saba, T.; Mustafa, L. Transforming Educational Insights: Strategic Integration of Federated Learning for Enhanced Prediction of Student Learning Outcomes. *J. Supercomput.* **2024**. [CrossRef]
86. Monteverde-Suárez, D.; González-Flores, P.; Santos-Solórzano, R.; García-Minjares, M.; Zavala-Sierra, I.; de la Luz, V.L.; Sánchez-Mendiola, M. Predicting Students' Academic Progress and Related Attributes in First-Year Medical Students: An Analysis with Artificial Neural Networks and Naïve Bayes. *BMC Med. Educ.* **2024**, *24*, 12909. [CrossRef]
87. Erdem, C.; Kaya, M. Socioeconomic Status and Wellbeing as Predictors of Students' Academic Achievement: Evidence from a Developing Country. *J. Psychol. Couns. Sch.* **2023**, *33*, 202–220. [CrossRef]
88. Dombkowski, R.; Sullivan, S.; Widenhoefer, T.; Buckland, A.; Almonroeder, T.G. Predicting First-Time National Physical Therapy Examination Performance for Graduates of an Entry-Level Physical Therapist Education Program. *J. Phys. Ther. Educ.* **2023**, *37*, 325–331. [CrossRef] [PubMed]
89. Liang, G.; Jiang, C.; Ping, Q.; Jiang, X. Academic Performance Prediction Associated with Synchronous Online Interactive Learning Behaviors Based on the Machine Learning Approach. *Interact. Learn. Environ.* **2023**, 1–16. [CrossRef]
90. Akçapinar, G.; Altun, A.; Aşkar, P. Modeling Students' Academic Performance Based on Their Interactions in an Online Learning Environment. *Elem. Educ. Online* **2015**, *14*, 815–824. [CrossRef]
91. Bergen, H.A.; Martin, G.; Roeger, L.; Allison, S. Perceived Academic Performance and Alcohol, Tobacco and Marijuana Use: Longitudinal Relationships in Young Community Adolescents. *Addict. Behav.* **2005**, *30*, 1563–1573. [CrossRef] [PubMed]
92. Teuber, M.; Leyhr, D.; Sudeck, G. Physical Activity Improves Stress Load, Recovery, and Academic Performance-Related Parameters among University Students: A Longitudinal Study on Daily Level. *BMC Public Health* **2024**, *24*, 598. [CrossRef] [PubMed]
93. Azpiazu, L.; Antonio-Aguirre, I.; Izar-de-la-Funte, I.; Fernández-Lasarte, O. School Adjustment in Adolescence Explained by Social Support, Resilience and Positive Affect. *Eur. J. Psychol. Educ.* **2024**, 0123456789. [CrossRef]
94. Odermatt, S.D.; Weidmann, R.; Schweizer, F.; Grob, A. Academic Performance through Multiple Lenses: Intelligence, Conscientiousness, and Achievement Striving Motivation as Differential Predictors of Objective and Subjective Measures of Academic Achievement in Two Studies of Adolescents. *J. Res. Pers.* **2024**, *109*, 104461. [CrossRef]
95. Jussila, J.J.; Pulakka, A.; Halonen, J.I.; Salo, P.; Allaouat, S.; Mikkonen, S.; Lanki, T. Are Active School Transport and Leisure-Time Physical Activity Associated with Performance and Wellbeing at Secondary School? A Population-Based Study. *Eur. J. Public Health* **2023**, *33*, 884–890. [CrossRef]
96. Ahmed, A.; Rashidi, M.Z. Predicting Perceived Academic Performance through Interplay of Self-Efficacy and Trait Emotional Intelligence. *Glob. Manag. J. Acad. Corp. Stud.* **2017**, *6*, 152–161.
97. Petrides, K.V. *Technical Manual for the Trait Emotional Intelligence Questionnaires (TEIQue)*, 1st ed.; London Psychometric Laboratory: London, UK, 2009. [CrossRef]
98. Petrides, K.V.; Furnham, A. Trait Emotional Intelligence: Psychometric Investigation with Reference to Established Trait Taxonomies. *Eur. J. Pers.* **2001**, *15*, 425–448. [CrossRef]
99. Jerusalem, M.; Schwarzer, R. Self-Efficacy as a Resource Factor in Stress Appraisal Processes. In *Self-Efficacy: Thought Control of Action*; Hemisphere Publishing Corp: Washington, DC, USA, 1992; pp. 195–213.
100. Torres, R.A.O.; Ortega-Dela Cruz, R.A. Remote Learning: Challenges and Opportunities for Educators and Students in the New Normal. *Anatol. J. Educ.* **2022**, *7*, 83–92. [CrossRef]
101. Corral, L.; Fronza, I. It's Great to Be Back: An Experience Report Comparing Course Satisfaction Surveys Before, During and After Pandemic. In Proceedings of the SIGITE 2022—Proceedings of the 23rd Annual Conference on Information Technology Education, Chicago, IL, USA, 21–24 September 2022; pp. 66–72. [CrossRef]
102. Al-Ansi, A.M.; Jaboob, M.; Garad, A.; Al-Ansi, A. Analyzing Augmented Reality (AR) and Virtual Reality (VR) Recent Development in Education. *Soc. Sci. Humanit. Open* **2023**, *8*, 100532. [CrossRef]
103. Hu Au, E.; Lee, J.J. Virtual Reality in Education: A Tool for Learning in the Experience Age. *Int. J. Innov. Educ.* **2017**, *4*, 215. [CrossRef]
104. Liddell, T.M.; Kruschke, J.K. Analyzing Ordinal Data with Metric Models: What Could Possibly Go Wrong? *J. Exp. Soc. Psychol.* **2018**, *79*, 328–348. [CrossRef]
105. Cohen, M.A.; Horowitz, T.S.; Wolfe, J.M. Auditory Recognition Memory Is Inferior to Visual Recognition Memory. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 6008–6010. [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.