International Journal of Data and Network Science 8 (2024) 2563–2580

Contents lists available at GrowingScience

International Journal of Data and Network Science

homepage: www.GrowingScience.com/ijds

Evaluation of factors associated with the adoption of ICT in education using machine learning

Holgado-Apaza Luis Alberto[a*], Aragon-Navarrete Ruth Nataly[a], Dioses-Córdova Ronald Román[b], Riva-Ruiz Raidith[c], Vidaurre-Rojas Pierre[c], Valles-Coral Miguel[c], Castellon-Apaza Danger David[a] and Quispe-Layme Marleny[a]

[a]*Universidad Nacional Amazónica de Madre de Dios, Puerto Maldonado 17001, Perú*
[b]*Consultores y Equipamiento D S.A.C, Puerto Maldonado 17001, Perú*
[c]*Universidad Nacional de San Martín, Tarapoto 22200, Perú*

C H R O N I C L E

Article history:
Received: January 6, 2024
Received in revised format: February 20, 2024
Accepted: May 2, 2024
Available online: May 2, 2024

Keywords:
ICT adoption
Technology acceptance
Feature selection
Educational predictive analytics
Educational technology integration

A B S T R A C T

Information and Communication Technologies (ICT) affect all aspects of our daily lives. Using them is considered a symbol of modernization and social advancement. The global expansion and interconnection of ICT offers a significant opportunity to promote the advancement of humanity, bridge the digital gap and promote the growth of societies built on knowledge. In this study, we analyzed and identified the most influential factors in the adoption of ICT in education from the data set called "Final Survey-Digital Inclusion Teachers" of the Plurinational State of Bolivia, which consists of 871 instances and 189 columns. We performed feature selection by carefully combining the results of three feature selection methods: filter (chi-square, ANO-VA and mutual information), wrapper (RFE) and intrinsic (Classification And Regression Trees, Random Forest, Gradient Boosting and XGBoost). The results demonstrated that a teacher's motivation for curricular planning that includes ICT, teaching experience and the institutional environment are key factors in the adoption of these technologies in education. Furthermore, we identified that the Random Forest algorithm is the most appropriate for analyzing and predicting the adoption of ICT in education, we affirmed this after this algorithm obtained the highest values in four of the six metrics evaluated: a sensitivity of 77.7%, an F1 Score of 77.9%, a Cohen's Kappa coefficient of 60.8% and a Jaccard Score of 64.3%. These results suggest that Random Forest is the most effective algorithm to analyze the factors related to the adoption of ICT in educational environments.

1. Introduction

Education stands out as one of the fundamental pillars in human life. Not only is it an essential need for the development of competencies, but it also facilitates personal recognition and progress (Alghamdi & Rahman, 2023; Selim & Rezk, 2023). Furthermore, education opens the doors to the discovery of the truth, reaffirming its importance in our society (Szymkowiak et al., 2021). Information and Communication Technologies (ICT) affect all aspects of our daily lives. Its use is considered a symbol of modernization and social advancement (Campos Cruz et al., 2018). The global expansion and interconnection of ICT offer a significant opportunity to promote the advancement of humanity, bridge the digital gap and promote the growth of societies built on knowledge-based societies (González-Zamar et al., 2020; Lawrence & Tar, 2018). In education, ICT plays a vital role in enabling students to explore concepts, engage in authentic and problem-based learning, and strengthen metacognitive skills. In addition, they allow the presentation of information in a diverse and dynamic way (Márquez et al., 2023). Advances in computer programs and multimedia tools available online suggest the possibility of an educational revolution, where the focus is oriented towards the student as the main protagonist of the learning process, while the role of the teacher evolves towards that of facilitator rather than mere transmitter of knowledge (Bansal, 2023; Lee et al., 2021) .

* Corresponding author.
E-mail address lholgado@unamad.edu.pe (H.-A. L. Alberto)

ISSN 2561-8156 (Online) - ISSN 2561-8148 (Print)

doi: 10.5267/j.ijdns.2024.5.002

The adoption of ICT in education offers greater possibilities to improve collaboration between teachers and students in a global digitalized context (Kler, 2014; Lawrence & Tar, 2018; Perienen, 2020). In this context, it is crucial for teachers to become familiar with, monitor and maintain a favorable attitude towards technology, in order to be able to effectively integrate technological innovations into the educational environment (Graham et al., 2020). Initiatives have been carried out to analyze the factors that influence the integration of ICT in education, as observed in the study presented by (Turgut & Aslan, 2021). This study focused on revealing the factors influencing ICT integration in academic environments in Turkey. The authors conducted an analysis of 60 studies using the meta-analysis method, leveraging various scholarly resources for their research. The results show that teachers' competence in ICT and pedagogy, students' proficiency in ICT, insufficient technological resources and assistance, scarcity of educational material, school administrators' attitudes, quality and deficiency in ICT training services impact the ICT integration in academic environments in Turkey. The study presented by Al-Mamary (2022) examined the aspects influencing the utilization of ICT in teaching as perceived by Yemeni educators. The authors administered a questionnaire to 120 teachers from both public and private schools, with representation from two schools in each one in Yemen. The findings indicate that the ease of accessing ICT infrastructure, receiving assistance from technical support teams, and having sufficient time and training for technology use are significant factors influencing teachers' utilization of technology in Yemen. In the study by Ifinedo et al. (2020) the authors investigated the factors influencing the technological integration of Nigerian teacher educators. The data were obtained with a questionnaire administered to 148 teacher educators in various departments. The technique used was partial least squares structural equation modeling. The findings reveal a statistical association between teaching experience and class size in the integration of technology among teacher educators in Nigeria. In the study by Habibi et al. (2020) explored the factors that affect ICT integration during teaching practices in initial teacher education programs at three Indonesian universities. The experiment's dataset was gathered through a survey conducted with 51 prospective teachers. The methodologies employed included internal case analysis and cross-case analysis. According to the results knowledge of ICT, perceived usefulness, perceived ease of use, support from leaders and support from colleagues constitute aspects that favor incorporating ICT into teaching methodologies.

The purpose of the study carried out by Madni et al. (2022) was to investigate the factors that impact the adoption of the Internet of Things (IoT) for e-learning in Higher Education Institutions. The research methodology involved a comprehensive review of relevant articles from various databases, such as Web of Science, Taylor & Francis, Springer, Scopus, Science Direct, Google Scholar, IEEE Explore and ACM Digital Library, covering publications between 2016 and 2021. The findings suggest that the adoption of IoT for e-learning is influenced by factors such as privacy concerns, infrastructure preparedness, financial limitations, usability, support from teachers, interaction, attitude, as well as network and data security. Other initiatives from the machine learning approach to address this problem are the following:

The research undertaken by Salman et al. (2023) sought to explore users' acceptance levels and attitudes toward incorporating Metaverse technology in higher education settings in Bahrain and Jordan. The methodology comprised two phases: structural equation modeling and the utilization of machine learning classification algorithms. The findings showed that users' attitudes toward utilizing this technology were significantly influenced by their perceptions of its ease of use and usefulness. In the study presented by Alhamad et al. (2021), they explored the adoption of Google Glass in the Gulf area in order to encourage its use in education, analyzing the relationship between the Technology Acceptance Model (TAM) and other factors. Data was collected from 420 questionnaires applied to students and teachers from various universities. The researchers employed partial least squares structural equation modeling along with machine learning models. The findings unveiled that the perceived usefulness and ease of use exerted a notable influence on the adoption of Google Glass. Although existing literature addresses the problem of ICT adoption in education, these studies, despite their value, have mainly focused on specific contexts and have used traditional methodological approaches. There is a notable absence of research that applies machine learning techniques to analyze and predict the factors that influence the global adoption of ICT in education. This disparity highlights the necessity for research that embraces innovative methodologies, such as machine learning, to attain a more profound comprehension of the factors linked with ICT adoption in education.

These are our contributions: We determine the factors that influence ICT adoption in education using an ensemble of feature selection methods. We identify that Random Forest is the most advisable algorithm for the analysis and prediction of the adoption of ICT in education.

We offer the academic community an innovative methodology to investigate the possible factors that could impact the adoption of ICT in education.

The subsequent sections of this paper will proceed as follows: Section 2 will delineate the materials and methods employed, Section 3 will present the results and discussions, and finally, Section 4 will encapsulate our conclusions.

2. Materials and Methods

Fig. 1 shows the framework we propose for the evaluation of factors associated with the ICT adoption in education using machine learning techniques. This process consists of four phases: (1) Obtaining "Encuesta Final-Profesores de Inclusión

Digital” data set (AGETIC, 2019). (2) Data cleaning and preprocessing. (3) Feature selection. (4) Model training and hyperparameter tuning, and (5) Evaluation and obtaining metrics.

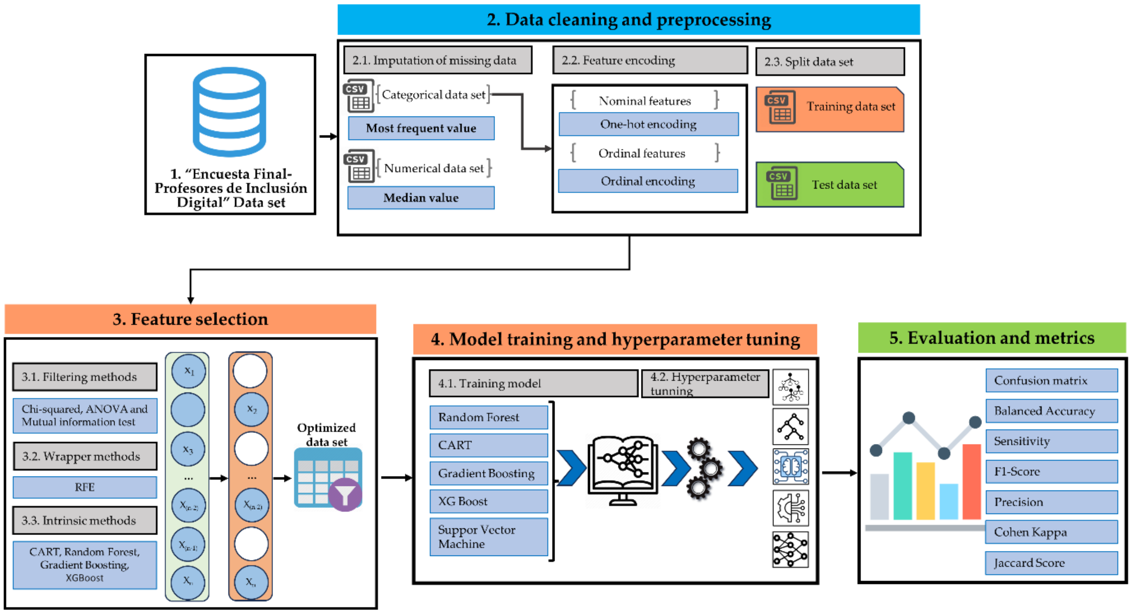

Fig. 1. Proposed framework for the evaluation of factors associated with the adoption of ICT in education using machine-learning techniques.

Below, we described in detail each of the phases of the proposed methodology, which constitute a systematic approach to evaluate the factors associated with the adoption of ICT in the educational field using machine-learning techniques.

2.1. Obtaining the teacher digital inclusion data set

We obtained the data set “Encuesta Final-Profesores de Inclusión Digital” from the open data platform of the Bolivia Plurinational State, focusing specifically on the education category (CTIC, 2023; Datos Abiertos, 2019). This data set, created by the Agency for Electronic Government and Information and Communication Technologies (AGETIC) in 2019, has as its main purpose the integration of free software technologies in teaching-learning methods in Public Educational Units (AGETIC, 2019). This resource, a detailed reflection of the situation of digital inclusion of teachers, provides a valuable and detailed perspective for our analysis. With a total of 871 rows and 189 columns, as shown in Table 1, we provide a general summary obtained after the initial exploration with the create_report() function of the dataprep.eda library (Peng et al., 2021) in Python applied to said data set.

Table 1
General statistics of the “Encuesta Final-Profesores de Inclusión Digital” data set

Attribute	Value obtained
Variables	189
Rows	871
All missing cells	35336
Missing cell (%)	21.5%
Duplicate rows	0
Duplicate rows (%)	0.0%
Variables types	Categorical: 184 Numerical: 5

2.2. Data cleaning and preprocessing

A crucial stage of the application of machine learning (ML) is data preprocessing (S. Zhang et al., 2003), which has been shown to significantly affect the performance of ML techniques (Huang et al., 2015; Lee & Chung, 2019; Pallathadka et al., 2023). In data science projects, data preprocessing typically requires 60 to 80% of the total time (Frye et al., 2021). The ratio may change depending on the intricacy of the data set, the initial quality of the data, and the specific nature of the addressed problem (Suprapto, 2024).

We began this phase by ensuring the quality of the data by filtering columns that contain, at most, 15% missing values, which resulted in 137 columns. Subsequently, we carried out the imputation of missing data using the most frequent value for categorical variables and the median for numerical variables, following recommended practices according to the literature review (Hussain & Khan, 2023; Rajendran et al., 2022; Srinivas & Rajendran, 2017).

Since the imputed data set consisted of 137 categorical columns and 5 numerical columns, we performed One-Hot Encoding for the 115 nominal variables, along with Ordinal Encoding for the 17 ordinal variables. This ensures that machine learning algorithms receive mainly numerical values as input, thus facilitating their processing and analysis (Choong & Lee, 2017; Seota et al., 2021). We performed these operations using the OrdinalEncoder class of the preprocessing module, present in the scikit-learn library (Pedregosa et al., 2011), and the get_dummies() function of the pandas library (The pandas development team, 2023) in Python, which resulted in increasing the dimension of the data set to 172 input columns and 1 output column.

Finally we have separated the data set into two groups: 80% for training and 20% for testing, a practice widely accepted in several studies with this approach (Celbiş et al., 2023; Segura et al., 2022; Yoo & Rho, 2020). Thus, we obtained a training dataset of 696 instances and 175 for testing.

2.3. Feature Selection

This process aims to find a subset of relevant and meaningful features to build a model (Barraza et al., 2019; Cang & Yu, 2012; Sulaiman & Labadin, 2015; Tang et al., 2014). For this task, filter and wrapper methods as well as intrinsic methods are used.

2.3.1. Applying filter methods

These methods evaluate features independently of the model by assigning scores (Albulayhi et al., 2022; Alsahaf et al., 2022; Pudjihartono et al., 2022). The m features that obtain the highest scores or those that exceed a threshold τ are selected with $m \in \mathrm{N}\ o\ \tau \in R$ (Bommert et al., 2020). Due to their low computational cost, univariate methods such as chi-square, ANOVA F, mutual information, Euclidean distance, Pearson correlation, Mann-Whitney test, t test, etc., have garnered greater attention (Chandrashekar & Sahin, 2014; Pudjihartono et al., 2022).

For this study, we employed the chi-square filter, ANOVA F test, and mutual information techniques. Notably, the ANOVA F test was utilized to select numerical features, while chi-square tests were applied for categorical variables. Mutual information was employed for both types of variables.

Chi-square filter scores, ANOVA F scores, and mutual information were obtained with the SelectKBest class from the feature_selection module of the scikit-learn Python library.

Eq. (1) shows the formula for calculating chi-square test scores.

$$X^2 = \sum_{i=1}^{r}\sum_{j=1}^{c}\frac{(O_{ij}-E_{ij})^2}{E_{ij}} \tag{1}$$

Having:
r represents the number of categories of the evaluated variable.
c represents the number of categories of the target variable.
O_{ij} observed values.
E_{ij} expected values.

Eq. (2) shows the formula for calculating ANOVA F scores.

$$F = \frac{\frac{SSB}{df_{between}}}{\frac{SSW}{df_{within}}} \tag{2}$$

Having:
SSB is the variability between groups and is defined by Eq. (3).

$$SSB = \sum_{i=1}^{k} n_i(\bar{X}_i - \bar{X}_{total})^2 \tag{3}$$

Additionally, k is the number of groups, n_i is the size of group i, $\bar{X}_i$ is the mean of group i, and $\bar{X}_{total}$ is the mean of all groups.

SSW is the variability within groups and is defined by Eq. (4).

$$SSW = \sum_{i=1}^{k} \sum_{j=1}^{n_i} (X_{ij} - \bar{X}_i)^2 \quad (4)$$

Here, X_{ij} is the value of individual j in group i, $\bar{X}_i$ is the mean of group i, and n_i is the size of the group i.

$df_{between}$ and df_{within} represent the degrees of freedom between groups and within groups respectively.

2.3.2. Applying the Recursive Feature Elimination (RFE) wrapper method

Wrapper methods assess the subset of potential features by evaluating the performance of a learning algorithm (Contreras et al., 2020; Pudjihartono et al., 2022; Remeseiro & Bolon-Canedo, 2019). Interaction with the classifier causes these methods to identify the best performing set of features for the selected algorithm (Li et al., 2016; Mandal et al., 2021; Wah et al., 2018). Nevertheless, these methods demand a high computational cost, because they need to train and evaluate the machine-learning model in each iteration of the feature selection process.

The FRE algorithm begins by building a model over the full set of predictors and calculating an importance score for each predictor. Minor predictors are then removed, the model is reconstructed, and importance scores are calculated again (Chen & Jeong, 2007).

We used the RFE class (Dhal & Azad, 2021; Miao & Niu, 2016) from the feature_selection module of the scikit-learn Python library for this process.

2.3.3. Application of intrinsic methods

Intrinsic or embedded methods incorporate the direct evaluation of features during the model training process, evaluating their importance based on their ability to predict a target variable (Brownlee et al., 2020). We apply algorithms like Classification and Regression Trees (CART) (Breiman et al., 2017; Trujillano et al., 2008), Random Forest (Breiman, 2001; Mukasheva et al., 2023), Gradient Boosting (Friedman, 2001) and XGBoost (Chen & Guestrin, 2016; Jin et al., 2024). These algorithms are not only effective in building predictive models, but also incorporate specific feature selection procedures.

2.3.4. Getting optimized data set

Obtaining an optimized data set is a crucial stage in the data analysis process. In this study, we have employed a thorough strategy that integrates findings from three distinct feature selection methods: filter methods, wrapper methods, and intrinsic methods. This approach, known in the data science literature as the “ensemble method for feature selection” (Pudjihartono et al., 2022), is based on the premise that the integration of various feature selection strategies can enhance the individual strengths of each method, thus providing a more robust and efficient data set (Bolón-Canedo et al., 2014). We combine the results of these methods carefully to form an optimized data set with twelve features where, eleven are predictors and one is our target variable. This approach is based on scientific evidence supporting the superiority of feature selection methods together, consistently demonstrating greater classification accuracy compared to the use of individual methods (Guney & Oztoprak, 2022; Pes, 2020; Spooner et al., 2023; Tsai & Sung, 2020; Wang et al., 2019). Table 2 shows the resulting optimized data set after executing this task. We show the questions that correspond to each of the columns in Table 3.

Table 2
Optimized data set

ID	i3_Sí	j11.SQ002	j11.SQ004	k4	e11	j11.SQ003	j11.SQ005	j11.SQ008	b2	c3	j11.SQ006	j11.SQ001
0	1	2	2	1	0.0	2	2	1	0.285	0	2	2
1	1	2	2	2	-1.0	2	2	2	0.214	0	2	2
2	1	2	2	2	0.0	2	2	1	0.285	1	2	1
…	…	…	…	…	…	…	…	…	…	…	…	…
868	1	2	2	2	-1.5	2	2	2	1.071	0	2	2
896	1	2	2	2	-1.0	2	2	2	0.357	0	2	2
870	0	1	1	2	-1.5	1	1	1	-0.357	0	1	1

Table 3
Description of selected columns

Variable	Description
i3_Sí	Have you ever used the KUAA to develop your classes?
j11.SQ002	Have you considered the development of the methodological moments (Practice-Theory-Assessment and Production), articulated to the use of KUAA or other ICT tools, for the development of the teaching-learning process?
j11.SQ004	When planning your lesson plans, have you ever started from the objective to define the use that could be made of KUAA or other ICT tools, which would be used by students to learn?
k4	In the development of the teaching-learning process, you use videos, presentations or infographics so that the student can consult and strengthen the knowledge of the advanced content in classes.
e11	How many days a week do you use the Internet?
j11.SQ003	You have considered the interests, motivations, and knowledge that your students had about the use of KUAA or other ICT tools, for the design of the teaching-learning process.
j11.SQ005	He thought of integrating KUAA or other ICT tools, to the teaching-learning process of students through the use of active methodologies, such as Problem Based Learning, Project Based Learning, Case Study Based Learning, Challenge Based Learning.
j11.SQ008	In order to integrate KUAA or other ICT tools in the classroom, I gathered or would gather information from other teachers.
b2	How many years of teaching experience do you have in your entire working life?
c3	How many computers (desktop or laptop) do you own?
j11.SQ006	In order to integrate KUAA or other ICT tools in the classroom, I did or would collect information from the director.
j11.SQ001	Have you incorporated KUAA or other ICT tools as a means to develop the teaching and learning process?

2.4. Model training and hyperparameter tuning

We trained our models, took advantage of the cloud computing capabilities of the Colab platform. In addition, we used a laptop with the following specifications: Intel® Core™ i5-1035G1 @ 1.00GHz processor, 8 processing cores, 16GB RAM and 64-bit Windows 10 Pro operating system. This configuration allowed us to efficiently carry out both the model training and hyperparameter optimization.

In order to start this procedure, we first confirm the class distribution in the training data set's target variable, which indicates the use of ICTs in educational settings. In our data collection, “never” is represented by value 0, “sometimes” by value 1, and “always” by value 2.

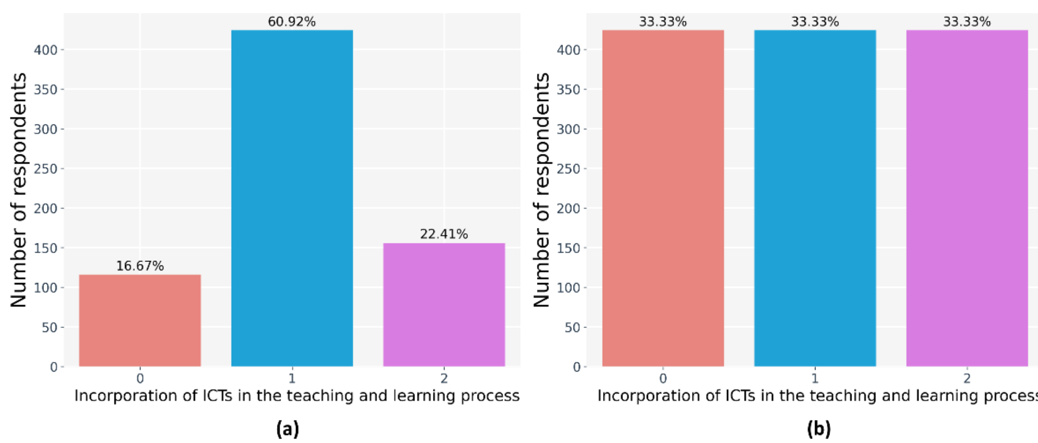

Fig. 2. Class distribution within the target variable: (a) Initial distribution; (b) Distribution post-oversampling

Fig. 2a shows this distribution, where we observe an imbalance ratio of 3.58:1 between the majority class (class 1) and the rarest class (class 0), indicating a moderate imbalance (Hasanin et al., 2019; He & Ma, 2013; Leevy et al., 2018). The presence of imbalance in the data could have a significant impact on predictive modeling (Batista et al., 2004; Zhang et al., 2020), presenting a tendency to skew the results in favor of the dominant class (Sambasivam & Opiyo, 2021). To address this challenge, we implemented the oversampling strategy, which involves adding instances to the minority classes until equilibrium with the majority class is achieved (Mohammed et al., 2020; Torres-Vásquez et al., 2021). This process allowed for an adjustment in the instance quantities per class, transitioning from 116 in class 0, 424 in class 1, and 156 in class 2, to 424 instances in each of the classes, as shown in Fig. 2b.

2.4.1. Model training

First, we used the default configurations as a baseline and trained the models using the CART, Random Forest, Gradient Boosting, XGBoost, and Support Vector Machine methods without adjusting any hyperparameters. On the basis of this baseline, we then aim to raise the assessment metrics' values.

2.4.2. Hyperparameter tuning

In machine learning, algorithms have critical hyperparameters that must be set before execution, and performance depends on their precise settings during training (Cruz Huacac, 2019; Padimi et al., 2023; Weerts et al., 2020). Many hyperameter optimization methods have been developed in the last 30 years (Bischl et al., 2023), ranging from traditional optimization methods such as gradient descent, decision theory-based approaches (grid search and random search), Bayesian optimization models, multi-fidelity optimization techniques and metaheuristic algorithms (Giselle Fernández-Godino, 2016; Yang & Shami, 2020).

Information theory-based approaches, such as grid search and random search, are widely successfully employed in machine learning projects (Abnoosian et al., 2023; Bergstra et al., 2012; Liashchynskyi & Liashchynskyi, 2019; Saputra et al., 2023; Shekar & Dagnew, 2019). While grid search is characterized by its exhaustiveness by searching for the optimal configuration in a fixed space of hyperparameters, random search selects combinations randomly in a defined space, thus achieving a reduction in execution time and optimal utilization of computational resources (Kaps et al., 2023; Saputra et al., 2023; Yang & Shami, 2020).

In this study, we implemented the random search strategy with 10-fold K-Fold stratified cross-validation and 100 repetitions on the training data set. The choice of this methodology was based on its outstanding performance, especially in contexts of hyperparameter optimization in higher dimensions, supported by previous research (Bischl et al., 2023; Elgeldawi et al., 2021; Villalobos-Arias et al., 2020). We show the hyperparameters, range of values, description, default values and optimized values for each model in Table 4.

Table 4
Hyperparameter configuration

Model	Hyperparameter	Range of Values	Description	Default Values	Optimal Values
CART	max_depth	[10:110] step 10+[None]	Max Tree Depth.	None	30
	criterion	[“gini”, “entropy”]	Metric used to assess the effectiveness of a division.	gini	gini
	min_samples_split	[2,5,10,15,20,25,30]	Minimum number of samples required for splitting on an internal node.	2	15
	min_samples_leaf	[1,2,4,6,8,10,12]	Minimum number of instances needed for a leaf node.	1	8
	max_features	["auto", "sqrt", "log2"]	Quantity of features to assess for optimal division.	None	auto
RF	n_estimators	[10:1000] step 100	Quantity of decision trees in the random forest.	100	230
	max_depth	[3,5,7,9,11,13,15,None]	Maximum tree depth.	None	None
	criterion	[“gini”, “entropy”]	Metric to assess the effectiveness of a split.	gini	entropy
	min_samples_split	[2,5,10]	Minimum number of instances needed to divide an internal node.	2	2
	min_samples_leaf	[1,2,4]	Minimum number of instances needed for a leaf node.	1	4
	max_features	[“auto”,“sqrt”,“log2”,None]	Number of features selected randomly without replacement for each split.	sqrt	log2
GB	n_estimators	[100:1000] step 100	Number of trees to train.	100	100
	max_depth	[3,5,8]	Maximum tree depth.	3	3
	criterion	[“friedman_mse”, “squared_error”]	Metric to assess the effectiveness of a split.	friedman_mse	friedman_mse
	min_samples_split	[500:595] step 5, [601:696] step 5, [702:797] step 5, [803:898] step 5, [904:994] step 5, 1000	Minimum number of instances needed to split an internal node.	2	914
	min_samples_leaf	[20,28,37,46,55,64,73,82,91,100]	Minimum number of instances needed to be a leaf node.	1	28
	max_features	[“log2”, “sqrt”]	Number of features to evaluate in the quest for the optimal split.	None	log2
	subsample	[0.5,0.618,0.8,0.85,0.9,0.95,1.0]	Fraction of samples to use to adjust individual base learners.	1.0	0.85
	loss	[“deviance”]	Loss function to optimize.	log_loss	deviance
	learning_rate	[0.01,0.025,0.05,0.075,0.1,0.15,0.2]	Learning rate.	0.1	0.1
XGB	n_estimators	[10,17,25,33,41 48,56,64 72,80]	Quantity of estimators to be created in the fitting process.	None	80
	max_depth	[3,5,7]	Maximum tree depth.	None	3
	colsample_bytree	[0.6,0.7,0.8,0.9]	Fraction of characteristics to consider for the construction of each tree.	None	0.8
	subsample	[0.6,0.7,0.8,0.9]	Fraction of observations to consider for the construction of each tree.	None	0.8
	learning_rate	[0.01,0.04,0.07,0.1]	Learning rate.	None	0.01
SVM	C	[0.1, 1, 10, 100, 1000]	Regularization parameter.	1.0	0.1
	gamma	[1,0.1,0.01,0.001,0.0001]	Coefficient for the kernel function.	scale	1
	kernel	[“linear”, “rbf”]	Kernel function type.	rbf	linear

2.5. Evaluation and obtaining metrics

We assess the models' performance using the test dataset, utilizing the following metrics: confusion matrix, balanced accuracy, sensitivity, F1-Score, precision, Cohen's kappa coefficient, and Jaccard index. Hereafter, we elaborate on each of these metrics.

2.5.1. Confusion matrix

It is a tabular representation that enables the visualization of a supervised learning algorithm's performance, offering a comprehensive breakdown of the counts of true positives (TP), false positives (FP), true negatives (TN), and false negatives (FN) (Fallucchi et al., 2020; Jain et al., 2020).

In Fig. 3, the confusion matrix corresponding to a multiclass classification problem with three classes (A, B, and C) is presented. In this matrix, the values AA, BB, and CC represent the true positives for classes A, B, and C, respectively, meaning the quantity of samples correctly classified as each class. On the other hand, AB represents the count of instances belonging to class A that were wrongly predicted as class B, in other words, the misclassifications. The quantity of false negatives for class A (FN_A) is determined by adding together the counts of AB and AC (FN_A = AB + AC), i.e., the sum of all class A samples wrongly classified as class B or C. To compute the false negative for any class within a row, the inaccuracies pertaining to that class or row are aggregated. Similarly, to determine the false positive for any predicted class within a column, all the inaccuracies of that column are summed. Therefore, the false positive for class A (FP_A) is obtained by adding BA and CA, represented as (FP_A=BA+CA) (Ballabio et al., 2018; Mahmudah et al., 2021; Tharwat, 2018; Wabang et al., 2022).

		Predicted Class			
		A	B	C	FN
True Class	A	AA	AB	AC	AB+AC
	B	BA	BB	BC	BA+BC
	C	CA	CB	CC	CA+CB
	FP	BA+CA	AB+CB	AC+BC	

Fig. 3. Confusion matrix for multiclass classification problems.

2.5.2. Balanced accuracy

This metric for evaluating the performance of a classification model considers the class imbalance in the data set. It is defined as the average of the recall obtained for each class, where 1 indicates the best value and 0 the worst value. Eq. (5) shows its calculation.

$$Balanced\ accuracy = \frac{Sensitivity + Specificity}{2} = \frac{1}{2}\left(\frac{TP}{TP + FN} + \frac{TN}{TN + FP}\right) \quad (5)$$

2.5.3. Sensitivity

Also referred to as true positive rate (TPR), hit rate, or recall, of a classifier, denotes the ratio of correctly identified positive samples to the total positive samples (Lovera & Cardinale, 2023; Tharwat, 2018). We show its calculation in Eq. (6).

$$Sensitivity = \frac{TP}{TP + FN} \quad (6)$$

2.5.4. Specificity

Recognized as the true negative rate (TNR) or inverse of recall, it represents the proportion of correctly classified negative samples relative to the total number of negative samples (Olabanjo et al., 2022). We show the calculation of this metric in Eq. (7).

$$Specificity = \frac{TN}{TN + FP} \quad (7)$$

2.5.5. Precision

It signifies the ratio of correctly classified positive samples to the total predicted positive samples (Holicza & Kiss, 2023). Eq. (8) shows the calculation of this metric.

$$Precision = \frac{TP}{TP + FP} \tag{8}$$

2.5.6. F1-Score

This metric provides a combined evaluation of the precision and recall of the classification model (Holicza & Kiss, 2023; Tharwat, 2018). The range of its value spans from zero to one, with higher scores suggesting superior classification performance. In Eq. (9) we show how to calculate this metric.

$$F1 - Score = 2\left(\frac{PrecisionxRecall}{Precision + Recall}\right) \tag{9}$$

2.5.7. Jaccard similarity

Measures the similarity between the real data set (ground truth) and the model predictions. Unlike other metrics, its calculation is not derived directly from the confusion matrix, but requires additional information about the original sets for its proper calculation. Eq. (10) allows us to obtain this value.

$$Jaccard\ similarity = \frac{|A \cap B|}{|A \cup B|} \tag{10}$$

Having:
A y B are two sets.
$|A \cap B|$ represents the extent of overlap between sets A and B.
$|A \cup B|$ represents the combined size of sets A and B.

2.5.7. Kappa coefficient

It is used to measure the level of agreement between two evaluators or classifiers, taking into account the possibility of agreement due to randomness. In machine learning it is used to evaluate the reliability of human annotations with the accuracy of machine learning models in classification tasks. Cohen's kappa coefficient fluctuates between 0 and 1, where 0 means random agreement and 1 means perfect agreement (Cerda & Villarroel Del, n.d.). A high Cohen's kappa value indicates better performance of a machine learning model. Eq. (11) shows the calculation of this metric.

$$Kappa\ coefficient = \frac{\Pr(a) - \Pr(e)}{1 - \Pr(e)} \tag{11}$$

Having:

Pr(a) denotes the observed agreement percentage.
Pr(e) signifies the probability of chance agreement between evaluators.

3. Results and Discussion

Fig. 4 showcases the most relevant variables for the ICT adoption in education, identified through the assembly method for feature selection. According to the feature selection results of the three applied methods, eleven variables stand out as the most influential in this context. These include experience with computer-assisted learning (KUAA) tools (i3_Si), consideration of methodological moments and ICT tools in the learning session (j11_SQ002), and integration of KUAA and ICT tools in session planning learning (j11_SQ004). Below are the variables: experience in using multimedia resources to reinforce learning in advanced content classes (k4), frequency of internet access (e11) and consideration of the interests, motivations and prior knowledge of the students. students about KUAA or other ICT tools (j11.SQ003). Subsequently, the variables are located: consideration to integrate the KUAA into active methodologies such as problem-based, project-based and straight learning (j11.SQ005), good relationship with colleagues to exchange information and experiences in incorporating ICT into their practice. pedagogical (j11.SQ008), work experience as a teacher (b2). Finally, the availability of technological equipment (c3) and the openness of managers to provide information for the integration of KUAA or other ICT tools in learning sessions (j11.SQ006).

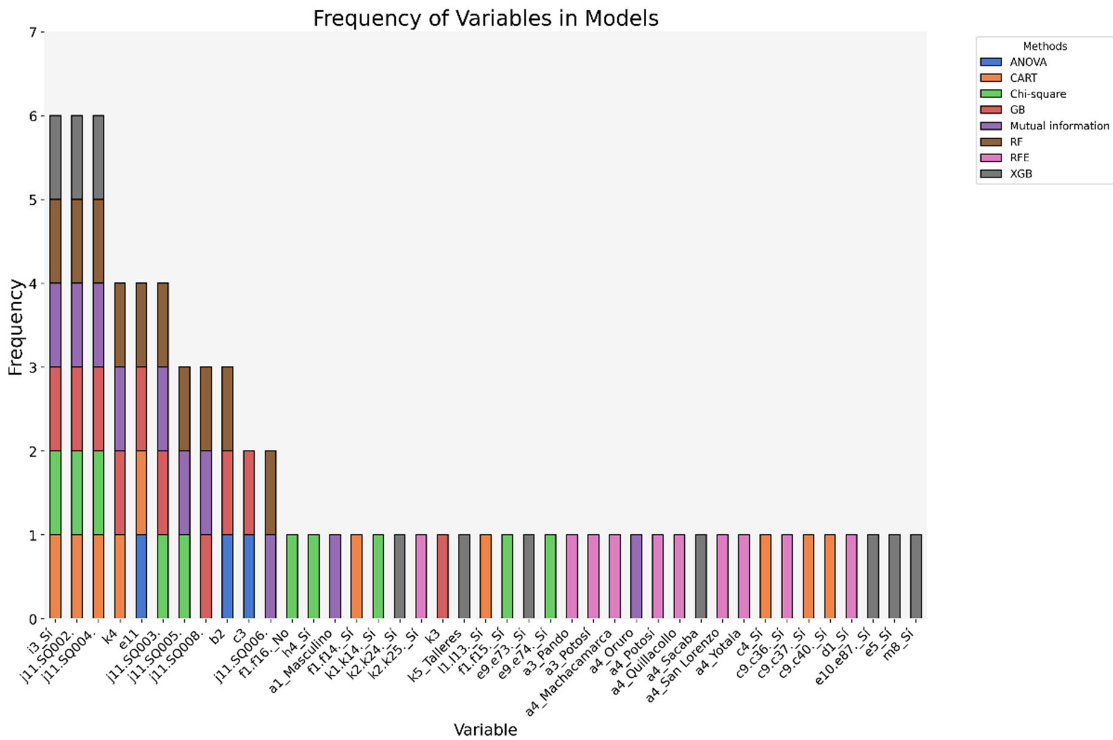

Fig. 4. Importance of variables in the adoption of Information and Communication Technologies in education.

The analysis of these findings allows us to identify three factors that significantly influence teachers' decision to adopt ICT in their pedagogical practice. The first of them is the teacher's motivation for curricular planning that includes ICT. The second is the teaching experience and the third factor is the institutional environment, understood as the set of institutional policies that foster a favorable climate for the feedback of experiences, as well as access to ICT (Irshad et al., 2023; Palacios Duarte & Saavedra García, 2018). On the one hand, the motivation of teachers to plan their curricular development incorporating ICT influences their adoption. This has been corroborated by the outstanding behavior of the variables: consideration of methodological moments and ICT tools in the learning session (j11_SQ002), integration of KUAA and ICT tools in the planning of learning sessions (j11_SQ004), the consideration of interests , motivations and prior knowledge of students about the KUAA or other ICT tools (j11.SQ003) and the consideration to integrate the KUAA into active methodologies such as problem-based, project-based and straight learning (j11.SQ005). The use of ICT can help students explore and learn authentically using various media (Bansal, 2023; González-Zamar et al., 2020). However, the incorporation of these technologies in education depends largely on the decision of teachers, therefore, their role is essential to develop students' digital competence (Baena-Morales et al., 2020). According to the results of this study, teachers' decision to adopt this technology is made from the moment of curricular planning. Additionally, the amalgamation of the methodological moments with ICT in the learning sessions and the interests of the students that teachers take into account, are aspects that can provide the actors responsible for education with useful contributions in curriculum development (Márquez et al., 2023). These results suggest the possibility of an educational revolution, with an approach oriented toward the student as the main protagonist of the learning process, and the role of the teacher as a facilitator instead of a mere transmitter of knowledge. This task also requires the support of an analysis of the students' interests, motivations and prior knowledge to guarantee that sustainable intervention proposals are as effective as possible (Dieste et al., 2019). On the other hand, it has been found that experience in education is a decisive factor for teachers to adopt ICT in their teaching work. These results are corroborated by the relevance that the variables have achieved: experience with computer-assisted learning tools (KUAA) (i3.Si), years of teaching work experience (b2) and experience in the use of multimedia resources to reinforce learning in advanced content classes (k4).

These results allow us to identify that the influence of teaching experience and experience in the use of technological tools come from prior knowledge and perceived knowledge in the use of ICT for its integration in education. Now, as ICT advances and its necessary incorporation into educational work, professional teacher training requires specific technological knowledge for the development of pedagogical practices and experiences incorporating ICT (Ifinedo et al., 2020). This is demonstrated by the outstanding behavior of the variables related to the teaching experience of the present study. Therefore, there is a need to emphasize the connection between teacher training and the development of their self-confidence to integrate technology into their pedagogical work (Paetsch et al., 2023). This experience reaches that related to a pandemic that could eventually contribute to the integration of ICT by teachers in the post-pandemic classroom (Paetsch & Drechsel, 2021). The last and not least important factor for the decision to adopt ICT in education comes from the human and technological institutional environment. The human aspect is given by the good relationship with colleagues and managers to exchange information and experiences in the incorporation of ICT in the teaching-learning processes (j11.SQ008 and j11.SQ006). The technological aspect corresponds to the technological infrastructure that has the provision of technological resources (i3_Si, c3) and access to Internet connectivity (e11). These results suggest the need for adequate availability of ICT in schools given its positive correlation with greater school efficiency (Agasisti et al., 2023). On the other hand, given that teachers' attitudes towards ICT depend on the context of the school's facilitating conditions, it is of utmost importance to analyze the efficiency of technology adoption in education from an institutional perspective, with the aim of providing appropriate ICT solutions to classrooms and educational environments. The good relationship between the director and his teachers can contribute to a better identification of results, with the aim of capturing suggestions and good practices (Cabellos et al., 2024). Equal importance is given to knowledge of privacy, infrastructure preparation, financial constraints, usability, teacher support, interaction, attitude and security of networks and data for the use of electronic educational platforms in higher education institutions (Madni et al., 2022). Therefore, easy access to ICT infrastructure, availability of support from the technical support team, availability of time and training for the use of technology and attitudes of school administrators are factors that impact the integration and use of technology (Al-Mamary, 2022; Turgut & Aslan, 2021).

Table 5 presents a detailed analysis of the performance of various machine learning models in the task of analyzing the adoption of ICT in the educational environment. These results are based on a comprehensive evaluation conducted through a stratified 10-fold cross-validation with 100 repetitions. We observed that the Random Forest model stands out as the most promising in terms of all the metrics evaluated, including Balanced Accuracy, Sensitivity, Precision, F1 Score, Cohen's Kappa coefficient, and Jaccard Score. Specifically, notable are the values achieved for the Balanced Accuracy metrics of 0.85, Sensitivity of 0.85, Precision of 0.87, and F1 Score of 0.85. These results suggest that this model possesses a remarkable ability to accurately identify the relevant factors for ICT adoption in the educational domain.

Table 5
Statistical overview of metrics derived from the training dataset

Model	Balanced Accuracy	Sensitivity	Precision	F1 Score	Cohen Kappa	Jaccard Score
Decision Trees-CART	0.76±0.11	0.76±0.11	0.78±0.13	0.75±0.12	0.64±0.17	0.63±0.15
Gradient Boosting	0.76±0.11	0.76±0.10	0.79±0.11	0.76±0.11	0.73±0.16	0.64±0.14
Random Forest	0.85±0.09	0.85±0.09	0.87±0.09	0.85±0.09	0.78±0.14	0.76±0.14
Support Vector Machine	0.73±0.12	0.73±0.12	0.74±0.15	0.71±0.14	0.60±0.18	0.60±0.16
XGBoost	0.78±0.11	0.78±0.11	0.81±0.10	0.77±0.11	0.66±0.16	0.66±0.14

These findings also suggest that the Random Forest algorithm may be especially valuable to researchers and practitioners interested in understanding and improving the integration of technology in educational settings. However, the other algorithms also show encouraging results, underscoring the diversity of approaches available in the field of machine learning to address complex educational problems. In Fig. 5, we show the distribution of metrics evaluated with a 10-fold, 100-replication stratified K-Fold cross-validation during the training stage. The graph highlights that the Random Forest algorithm shows a higher median in all the evaluated metrics compared to the other algorithms. Although some outliers are observed in this algorithm, its performance is still robust, as the interquartile variability of the data, indicated by the length of the green box, is smaller compared to other algorithms. Furthermore, the small size of the whiskers suggests that the dispersion of the data beyond the interquartile range is limited, supporting the consistency of the results of the Random Forest algorithm.

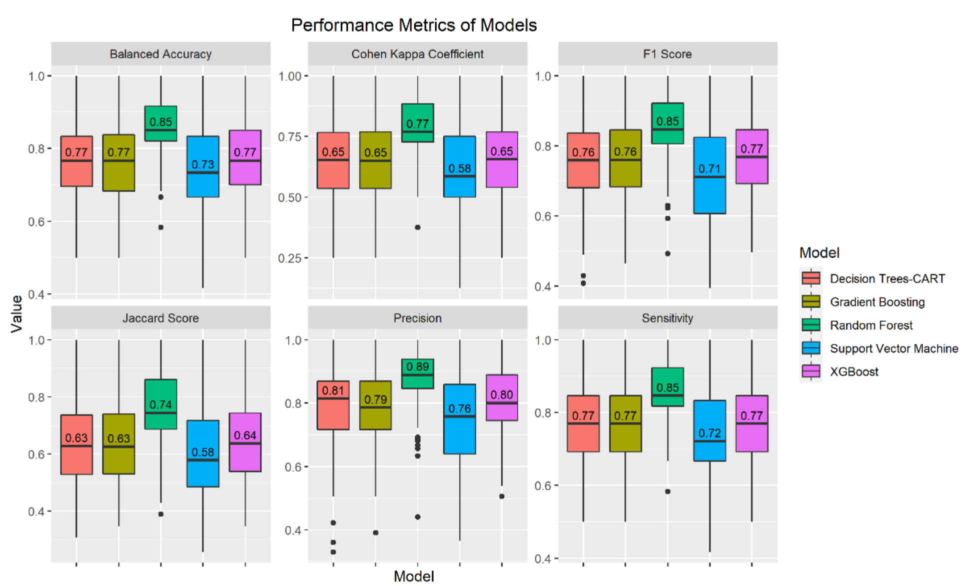

Fig. 5. Distribution of the evaluation metrics of machine learning models in the adoption of Information and Communication Technologies in Education in the training data set.

Fig. 6 displays the confusion matrices of the models assessed on the test data set concerning the adoption of ICT in the educational realm. We observe that in the quadrant of true positives for class 0, which represents teachers who have never adopted ICT in their learning sessions, the five algorithms evaluated manage to classify approximately 23 instances correctly.

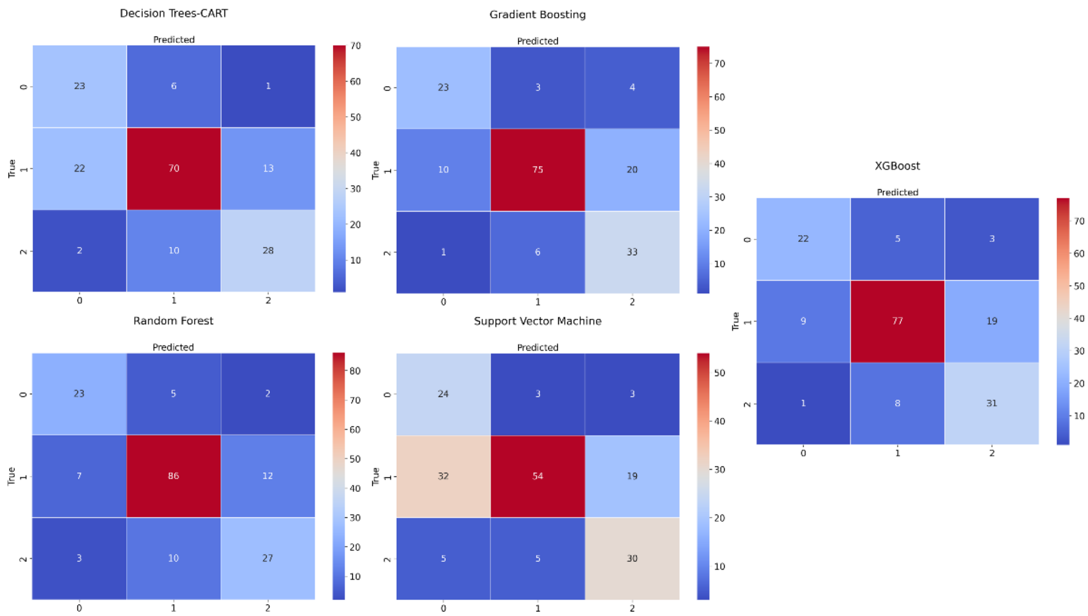

Fig. 6. Confusion matrices of the models analyzed in the adoption of Information and Communication Technologies in Education.

On the other hand, in the quadrant of true positives for class 1, which represents teachers who have sometimes adopted ICT in their learning sessions, we observe that the Random Forest algorithm manages to correctly classify 86 instances, this being the highest in this category. Finally, for the case of true positives in class 2, which represents teachers who always adopt ICT in their learning sessions, it is observed that the Gradient Boosting algorithm manages to classify 33 instances correctly, which indicates superior performance in this category. For a more robust and reliable evaluation of machine learning models, we compute various metrics using confusion matrices. The results are presented in Fig. 7, where it is observed that the Random Forest model shows outstanding performance by obtaining the best values in four of the six metrics evaluated. Specifically, it achieves a sensitivity of 77.7%, an F1 Score of 77.9%, a Cohen's Kappa coefficient of 60.8% and a Jaccard Score of 64.3%.

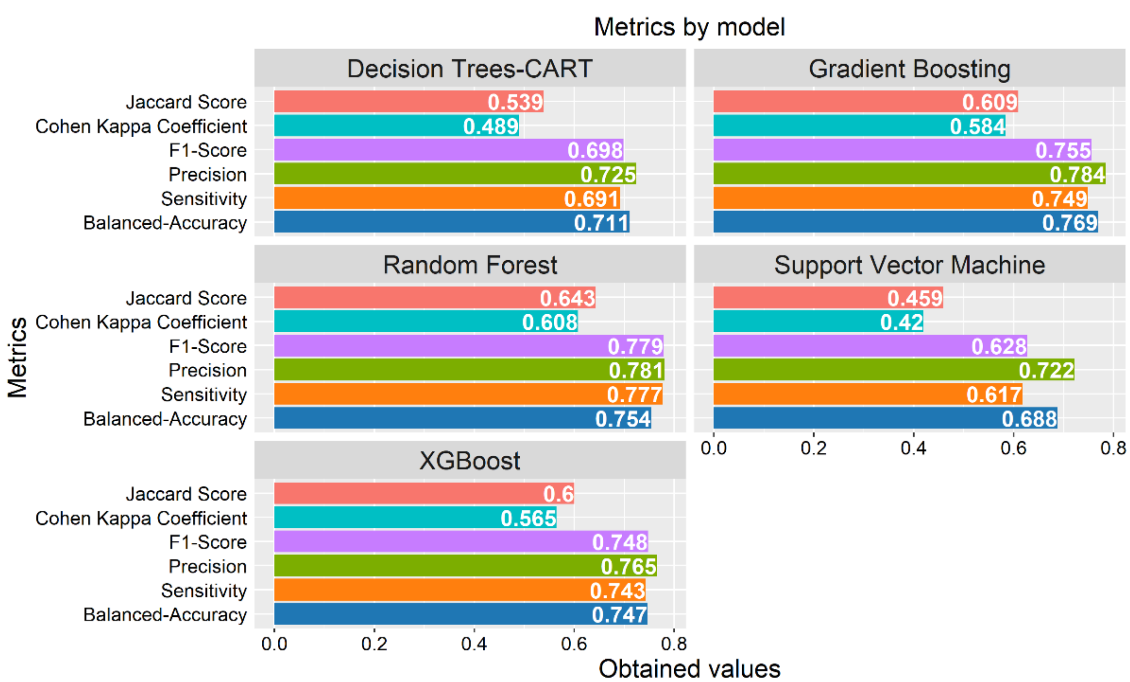

Fig. 7. Results of machine learning model evaluation metrics on the test data set.

In precision, the Random Forest algorithm attains a value of 78.1%, closely matching the 78.4% achieved by the Gradient Boosting algorithm. On the other hand, regarding the Balanced Accuracy metric, the Random Forest algorithm reaches a value of 75.4%, slightly lower than the 76.9% obtained by the Gradient Boosting algorithm. These findings suggest that the Random Forest algorithm is a solid and effective option to analyze factors related to the adoption of ICT in educational environments. However, it is also observed that other algorithms, such as Gradient Boosting and XGBoost, show promising results and could be considered as viable alternatives in this context.

4. Conclusion

In this study, we analyze and identify the most influential factors in the adoption of ICT in education using the data set "Final Survey-Digital Inclusion Teachers" of the Plurinational State of Bolivia, which consists of 871 instances and 189 columns. To enhance the credibility of our findings, we meticulously integrate the outcomes of three feature selection methods: filter (chi-square, ANOVA and mutual information), wrapper (RFE) and intrinsic (CART, Random Forest, Gradient Boosting and XGBoost). Our findings demonstrate that teacher motivation for curricular planning that includes ICT, teaching experience and the institutional environment, the latter understood as the set of institutional policies that foster a favorable climate for feedback on experiences, as well as access to ICT are key factors in the adoption of these technologies in education. Furthermore, we have identified that the Random Forest algorithm is the most suitable for analyzing and predicting the adoption of ICT in education, since it has obtained the highest values in four of the six metrics evaluated: a sensitivity of 77.7%, an F1 Score of 77.9%, a Cohen's Kappa coefficient of 60.8% and a Jaccard Score of 64.3%. These results suggest that Random Forest is the most effective algorithm to analyze the factors associated to the adoption of ICT in educational environments. We must clarify that our research focused solely on the analysis of data from a single source, specifically the "Final Survey-Digital Inclusion Teachers" of the Plurinational State of Bolivia. Furthermore, although a comprehensive feature selection approach incorporating different methods was used, some relevant factors may not have been included in the analysis. We hope that this study lays the foundation for future research that applies the machine learning approach to analyze the factors related to the adoption of ICT in educational settings. Given the paucity of similar studies in the literature, our findings offer a significant contribution to the field.

References

Abnoosian, K., Farnoosh, R., & Behzadi, M. H. (2023). Prediction of diabetes disease using an ensemble of machine learning multi-classifier models. *BMC Bioinformatics*, *24*(1), 1–24. https://doi.org/10.1186/S12859-023-05465-Z/FIGURES/7

Agasisti, T., Antequera, G., & Delprato, M. (2023). Technological resources, ICT use and schools efficiency in Latin America – Insights from OECD PISA 2018. *International Journal of Educational Development*, *99*, 102757. https://doi.org/10.1016/J.IJEDUDEV.2023.102757

AGETIC. (2019). *Encuesta Final-Profesores de Inclusión Digital - Conjuntos de datos - Datos Abiertos Bolivia*.

Al-Mamary, Y. H. S. (2022). Examining the factors affecting the use of ICT in teaching in Yemeni schools. *Journal of Public Affairs*, *22*(1), e2330. https://doi.org/10.1002/PA.2330

Albulayhi, K., Al-Haija, Q. A., Alsuhibany, S. A., Jillepalli, A. A., Ashrafuzzaman, M., & Sheldon, F. T. (2022). IoT Intrusion Detection Using Machine Learning with a Novel High Performing Feature Selection Method. *Applied Sciences 2022, Vol. 12, Page 5015*, *12*(10), 5015. https://doi.org/10.3390/APP12105015

Alghamdi, A. S., & Rahman, A. (2023). Data Mining Approach to Predict Success of Secondary School Students: A Saudi Arabian Case Study. *Education Sciences 2023, Vol. 13, Page 293*, *13*(3), 293. https://doi.org/10.3390/EDUCSCI13030293

Alhamad, A. Q. M., Akour, I., Alshurideh, M., Al-Hamad, A. Q., Kurdi, B. Al, & Alzoubi, H. (2021). Predicting the intention to use google glass: A comparative approach using machine learning models and PLS-SEM. *Volume 5, Issue 3, Pages 311 - 320*, *5*(3), 311–320. https://doi.org/10.5267/j.ijdns.2021.6.002

Alsahaf, A., Petkov, N., Shenoy, V., & Azzopardi, G. (2022). A framework for feature selection through boosting. *Expert Systems with Applications*, *187*, 115895. https://doi.org/10.1016/J.ESWA.2021.115895

Baena-Morales, S., Martinez-Roig, R., & Hernádez-Amorós, M. J. (2020). Sustainability and Educational Technology—A Description of the Teaching Self-Concept. *Sustainability, 12*(24), 10309. https://doi.org/10.3390/SU122410309

Ballabio, D., Grisoni, F., & Todeschini, R. (2018). Multivariate comparison of classification performance measures. *Chemometrics and Intelligent Laboratory Systems*, *174*, 33–44. https://doi.org/10.1016/J.CHEMOLAB.2017.12.004

Bansal, M. (2023). Use of information and communication technology (ICT) in education system. *International Journal of Multidisciplinary Trends*, *5*(3), 04–06. https://www.multisubjectjournal.com/archives/2023.v5.i3.A.263

Barraza, N., Moro, S., Ferreyra, M., & de la Peña, A. (2019). Mutual information and sensitivity analysis for feature selection in customer targeting: A comparative study. *Journal of Information Science*, *45*(1), 53–67. https://doi.org/10.1177/0165551518770967/ASSET/IMAGES/LARGE/10.1177_0165551518770967-FIG4.JPEG

Batista, G. E. A. P. A., Prati, R. C., & Monard, M. C. (2004). A study of the behavior of several methods for balancing machine learning training data. *ACM SIGKDD Explorations Newsletter*, *6*(1), 20–29. https://doi.org/10.1145/1007730.1007735

Bergstra, J., Ca, J. B., & Ca, Y. B. (2012). Random Search for Hyper-Parameter Optimization Yoshua Bengio. *Journal of Machine Learning Research*, *13*, 281–305. http://scikit-learn.sourceforge.net.

Bischl, B., Binder, M., Lang, M., Pielok, T., Richter, J., Coors, S., Thomas, J., Ullmann, T., Becker, M., Boulesteix, A. L., Deng, D., & Lindauer, M. (2023). Hyperparameter optimization: Foundations, algorithms, best practices, and open challenges. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *13*(2), e1484. https://doi.org/10.1002/WIDM.1484

Bolón-Canedo, V., Sánchez-Maroño, N., Alonso-Betanzos, A., Benítez, J. M., & Herrera, F. (2014). A review of microarray datasets and applied feature selection methods. *Information Sciences*, *282*, 111–135. https://doi.org/10.1016/J.INS.2014.05.042

Bommert, A., Sun, X., Bischl, B., Rahnenführer, J., & Lang, M. (2020). Benchmark for filter methods for feature selection in high-dimensional classification data. *Computational Statistics & Data Analysis*, *143*, 106839. https://doi.org/10.1016/J.CSDA.2019.106839

Breiman, L. (2001). Random Forests. *Machine Learning*, *45*(1), 5–32. https://doi.org/10.1023/A:1010933404324

Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (2017). Classification and regression trees. *Classification and Regression Trees*, 1–358. https://doi.org/10.1201/9781315139470/CLASSIFICATION-REGRESSION-TREES-LEO-BREIMAN

Brownlee, J., Sanderson, M., Koshy, A., Cheremskoy, A., & Halfyard, J. (2020). *Machine Learning Mastery With Python: Data Cleaning, Feature Selection, and Data Transforms in Python*.

Cabellos, B., Siddiq, F., & Scherer, R. (2024). The moderating role of school facilitating conditions and attitudes towards ICT on teachers' ICT use and emphasis on developing students' digital skills. *Computers in Human Behavior*, *150*, 107994. https://doi.org/10.1016/J.CHB.2023.107994

Campos Cruz, H., Ramírez Sánchez, M. Y., Campos Cruz, H., & Ramírez Sánchez, M. Y. (2018). Las TIC en los procesos educativos de un centro público de investigación. *Apertura (Guadalajara, Jal.)*, *10*(1), 56–70. https://doi.org/10.32870/AP.V10N1.1160

Cang, S., & Yu, H. (2012). Mutual information based input feature selection for classification problems. *Decision Support Systems*, *54*(1), 691–698. https://doi.org/10.1016/J.DSS.2012.08.014

Celbiş, M. G., Wong, P. H., Kourtit, K., & Nijkamp, P. (2023). Job Satisfaction and the 'Great Resignation': An Exploratory Machine Learning Analysis. *Social Indicators Research*, *170*(3), 1097–1118. https://doi.org/10.1007/S11205-023-03233-3/FIGURES/6

Cerda, J. L., & Villarroel Del, L. P. (n.d.). *Evaluación de la concordancia inter-observador en investigación pediátrica: Coeficiente de Kappa*.

Chandrashekar, G., & Sahin, F. (2014). A survey on feature selection methods. *Computers & Electrical Engineering*, *40*(1), 16–28. https://doi.org/10.1016/J.COMPELECENG.2013.11.024

Chen, T., & Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785–794. https://doi.org/10.1145/2939672

Chen, X., & Jeong, J. C. (2007). Enhanced recursive feature elimination. *Sixth International Conference on Machine Learning and Applications (ICMLA 2007)*, 429–435. https://doi.org/10.1109/ICMLA.2007.35

Choong, A. C. H., & Lee, N. K. (2017). Evaluation of convolutionary neural networks modeling of DNA sequences using ordinal versus one-hot encoding method. *1st International Conference on Computer and Drone Applications: Ethical Integration of Computer and Drone Technology for Humanity Sustainability, IConDA 2017*, *2018-Janua*, 60–65. https://doi.org/10.1109/ICONDA.2017.8270400

Contreras, L. E., Fuentes, H. J., Rodríguez, J. I., Contreras, L. E., Fuentes, H. J., & Rodríguez, J. I. (2020). Predicción del rendimiento académico como indicador de éxito/fracaso de los estudiantes de ingeniería, mediante aprendizaje automático. *Formación Universitaria*, *13*(5), 233–246. https://doi.org/10.4067/S0718-50062020000500233

Cruz Huacac, R. D. (2019). *Uso de Máquina de Soporte Vectorial para Predicción de Resistencia a la Compresión de Concreto* [Universidad Nacional de San Agustín de Arequipa]. http://repositorio.unsa.edu.pe/handle/UNSA/9092

CTIC. (2023). *CTIC – Consejo de Tecnologías e Información del Estado Plurinacional de Bolivia*.

Datos Abiertos. (2019). *Bienvenida - Datos Abiertos Bolivia*.

Dhal, P., & Azad, C. (2021). A comprehensive survey on feature selection in the various fields of machine learning. *Applied Intelligence*, *52*(4), 4543–4581. https://doi.org/10.1007/S10489-021-02550-9

Dieste, B., Coma, T., & Blasco-Serrano, A. C. (2019). Inclusión de los Objetivos de Desarrollo Sostenible en el Currículum de Educación Primaria y Secundaria en Escuelas Rurales de Zaragoza. *Revista Internacional de Educación Para La Justicia Social*, *8*(1), 97–115. https://doi.org/10.15366/RIEJS2019.8.1.006

Elgeldawi, E., Sayed, A., Galal, A. R., & Zaki, A. M. (2021). Hyperparameter Tuning for Machine Learning Algorithms Used for Arabic Sentiment Analysis. *Informatics, 8*(4), 79. https://doi.org/10.3390/INFORMATICS8040079

Fallucchi, F., Coladangelo, M., Giuliano, R., & De Luca, E. W. (2020). Predicting employee attrition using machine learning techniques. *Computers*, *9*(4), 1–17. https://doi.org/10.3390/computers9040086

Friedman, J. (2001). Greedy Function Approximation: A Gradient Boosting Machine on JSTOR. *The Annals of Statistics*, *29*(5), 1189–1232. https://www.jstor.org/stable/2699986

Frye, M., Mohren, J., & Schmitt, R. H. (2021). Benchmarking of Data Preprocessing Methods for Machine Learning-Applications in Production. *Procedia CIRP*, *104*, 50–55. https://doi.org/10.1016/J.PROCIR.2021.11.009

Giselle Fernández-Godino, P. M. (2016). *Review of multi-fidelity models*. https://arxiv.org/abs/1609.07196v4

González-Zamar, M. D., Abad-Segura, E., López-Meneses, E., & Gómez-Galán, J. (2020). Managing ICT for Sustainable Education: Research Analysis in the Context of Higher Education. *Sustainability, 12*(19), 8254. https://doi.org/10.3390/SU12198254

Graham, M. A., Stols, G., & Kapp, R. (2020). Teacher Practice and Integration of ICT: Why Are or Aren't South African Teachers Using ICTs in Their Classrooms. *International Journal of Instruction*, *13*(2), 749–766. https://doi.org/10.29333/iji.2020.13251a

Guney, H., & Oztoprak, H. (2022). A robust ensemble feature selection technique for high-dimensional datasets based on minimum weight threshold method. *Computational Intelligence*, *38*(5), 1616–1658. https://doi.org/10.1111/COIN.12524

Habibi, A., Razak, R. A., Yusop, F. D., Mukminin, A., & Yaqin, L. N. (2020). Factors Affecting ICT Integration During Teaching Practices: A Multiple Case Study of Three Indonesian Universities. *The Qualitative Report*, *25*(5), 1127–1144. https://doi.org/10.46743/2160-3715/2020.4150

Hasanin, T., Khoshgoftaar, T. M., Leevy, J. L., & Bauder, R. A. (2019). Severely imbalanced Big Data challenges: investigating data sampling approaches. *Journal of Big Data*, *6*(1), 1–25. https://doi.org/10.1186/S40537-019-0274-4/FIGURES/3

He, H., & Ma, Y. (2013). Imbalanced learning: Foundations, algorithms, and applications. In *Imbalanced Learning: Foundations, Algorithms, and Applications*. wiley. https://doi.org/10.1002/9781118646106

Holicza, B., & Kiss, A. (2023). Predicting and Comparing Students' Online and Offline Academic Performance Using Machine Learning Algorithms. *Behavioral Sciences, 13*(4), 289. https://doi.org/10.3390/BS13040289

Huang, J., Li, Y. F., & Xie, M. (2015). An empirical analysis of data preprocessing for machine learning-based software cost estimation. *Information and Software Technology*, *67*, 108–127. https://doi.org/10.1016/J.INFSOF.2015.07.004

Hussain, S., & Khan, M. Q. (2023). Student-Performulator: Predicting Students' Academic Performance at Secondary and Intermediate Level Using Machine Learning. *Annals of Data Science*, *10*(3), 637–655. https://doi.org/10.1007/S40745-021-00341-0/FIGURES/4

Ifinedo, E., Rikala, J., & Hämäläinen, T. (2020). Factors affecting Nigerian teacher educators' technology integration: Considering characteristics, knowledge constructs, ICT practices and beliefs. *Computers & Education*, *146*, 103760. https://doi.org/10.1016/J.COMPEDU.2019.103760

Irshad, M., Qureshi, M. A., Saraih, U. N., & Ahmad, S. F. (2023). Impact of institutional climate on the student's engagement and learning outcomes in private sector universities of Karachi. *International Journal of Management in Education*, *17*(3), 297–322. https://doi.org/10.1504/IJMIE.2023.130674

Jain, P. K., Jain, M., & Pamula, R. (2020). Explaining and predicting employees' attrition: a machine learning approach. *SN Applied Sciences*, *2*(4), 1–11. https://doi.org/10.1007/S42452-020-2519-4/TABLES/5

Jin, Y., Xu, S., Shao, Z., Luo, X., Wang, Y., Yu, Y., & Wang, Y. (2024). Discovery of depression-associated factors among childhood trauma victims from a large sample size: Using machine learning and network analysis. *Journal of Affective Disorders*, *345*, 300–310. https://doi.org/10.1016/J.JAD.2023.10.101

Kaps, A., Lehrer, T., Lepenies, I., Wagner, M., & Duddeck, F. (2023). Multi-fidelity optimization of metal sheets concerning manufacturability in deep-drawing processes. *Structural and Multidisciplinary Optimization*, *66*(8), 1–16. https://doi.org/10.1007/S00158-023-03631-8/FIGURES/13

Kler, S. (2014). ICT Integration in Teaching and Learning: Empowerment of Education with Technology. *Issues and Ideas in Education*, *2*(2), 255–271. https://doi.org/10.15415/IIE.2014.22019

Lawrence, J. E., & Tar, U. A. (2018). Factors that influence teachers' adoption and integration of ICT in teaching/learning process. *Educational Media International*, *55*(1), 79–105. https://doi.org/10.1080/09523987.2018.1439712

Lee, S., & Chung, J. Y. (2019). The Machine Learning-Based Dropout Early Warning System for Improving the Performance of Dropout Prediction. *Applied Sciences 2019, Vol. 9, Page 3093*, *9*(15), 3093. https://doi.org/10.3390/APP9153093

Lee, Y. C., Malcein, L. A., & Kim, S. C. (2021). Information and communications technology (ICT) usage during COVID-19: Motivating factors and implications. *International Journal of Environmental Research and Public Health*, *18*(7). https://doi.org/10.3390/IJERPH18073571/S1

Leevy, J. L., Khoshgoftaar, T. M., Bauder, R. A., & Seliya, N. (2018). A survey on addressing high-class imbalance in big data. *Journal of Big Data*, *5*(1), 1–30. https://doi.org/10.1186/S40537-018-0151-6/TABLES/5

Li, J., Cheng, K., Wang, S., Morstatter, F., Trevino, R. P., Tang, J., & Liu, H. (2016). Feature Selection: A Data Perspective. *ACM Computing Surveys*, *50*(6). https://doi.org/10.1145/3136625

Liashchynskyi, P., & Liashchynskyi, P. (2019). *Grid Search, Random Search, Genetic Algorithm: A Big Comparison for NAS*. *1*. https://arxiv.org/abs/1912.06059v1

Lovera, F. A., & Cardinale, Y. (2023). Análisis de sentimientos en Twitter: Un estudio comparativo. *Revista Científica de Sistemas e Informática*, *3*(1), e418–e418. https://doi.org/10.51252/RCSI.V3I1.418

Madni, S. H. H., Ali, J., Husnain, H. A., Masum, M. H., Mustafa, S., Shuja, J., Maray, M., & Hosseini, S. (2022). Factors Influencing the Adoption of IoT for E-Learning in Higher Educational Institutes in Developing Countries. *Frontiers in Psychology*, *13*, 915596. https://doi.org/10.3389/FPSYG.2022.915596/BIBTEX

Mahmudah, K. R., Purnama, B., Indriani, F., & Satou, K. (2021). Machine learning algorithms for predicting chronic obstructive pulmonary disease from gene expression data with class imbalance. *BIOINFORMATICS 2021 - 12th International Conference on Bioinformatics Models, Methods and Algorithms; Part of the 14th International Joint Conference on Biomedical Engineering Systems and Technologies, BIOSTEC 2021*, 148–153. https://doi.org/10.5220/0010316501480153

Mandal, M., Singh, P. K., Ijaz, M. F., Shafi, J., & Sarkar, R. (2021). A Tri-Stage Wrapper-Filter Feature Selection Framework for Disease Classification. *Sensors*, *21*(16), 5571. https://doi.org/10.3390/S21165571

Márquez, S., Jesús, J., Linares, G., Del Carmen Pérez-Fuentes, M., & Makrakis, V. (2023). Using the DREAM Methodology for Course Assessment in the Field of ICT-Enabled Education for Sustainability. *European Journal of Investigation in Health, Psychology and Education*, *13*(7), 1378–1391. https://doi.org/10.3390/EJIHPE13070100

Miao, J., & Niu, L. (2016). A Survey on Feature Selection. *Procedia Computer Science*, *91*, 919–926. https://doi.org/10.1016/J.PROCS.2016.07.111

Mohammed, R., Rawashdeh, J., & Abdullah, M. (2020). Machine Learning with Oversampling and Undersampling Techniques: Overview Study and Experimental Results. *2020 11th International Conference on Information and Communication Systems, ICICS 2020*, 243–248. https://doi.org/10.1109/ICICS49469.2020.239556

Mukasheva, M., Mukhiyadin, A., Makhazhanova, U., & Serikbayeva, S. (2023). The Behaviour of the Ensemble Learning Model in Analysing Educational Data on COVID-19. *13*(12), 1868–1878. https://doi.org/10.18178/ijiet.2023.13.12.2000

Olabanjo, O. A., Wusu, A. S., & Manuel, M. (2022). A machine learning prediction of academic performance of secondary school students using radial basis function neural network. *Trends in Neuroscience and Education*, *29*, 100190. https://doi.org/10.1016/J.TINE.2022.100190

Padimi, V., Telu, V. S., & Ningombam, D. D. (2023). Performance analysis and comparison of various machine learning algorithms for early stroke prediction. *ETRI Journal*, *45*(6), 1007–1021. https://doi.org/10.4218/ETRIJ.2022-0271

Paetsch, J., & Drechsel, B. (2021). Factors Influencing Pre-service Teachers' Intention to Use Digital Learning Materials: A Study Conducted During the COVID-19 Pandemic in Germany. *Frontiers in Psychology*, *12*, 733830. https://doi.org/10.3389/FPSYG.2021.733830/BIBTEX

Paetsch, J., Franz, S., & Wolter, I. (2023). Changes in early career teachers' technology use for teaching: The roles of teacher self-efficacy, ICT literacy, and experience during COVID-19 school closure. *Teaching and Teacher Education*, *135*, 104318. https://doi.org/10.1016/J.TATE.2023.104318

Palacios Duarte, P. D., & Saavedra García, M. L. (2018). Institutional R&D environment and its influence on employment and sales in Mexican manufacturing SMEs. *Revista Finanzas y Politica Economica*, *10*(1), 111–133. https://doi.org/10.14718/REVFINANZPOLITECON.2018.10.1.4

Pallathadka, H., Wenda, A., Ramirez-Asís, E., Asís-López, M., Flores-Albornoz, J., & Phasinam, K. (2023). Classification and prediction of student performance data using various machine learning algorithms. *Materials Today: Proceedings*, *80*, 3782–3785. https://doi.org/10.1016/J.MATPR.2021.07.382

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, E. (2011). Scikit-learn: Machine Learning in {P}ython. *Journal of Machine Learning Research*, *12*, 2825–2830.

Peng, J., Wu, W., Lockhart, B., Bian, S., Yan, J. N., Xu, L., Chi, Z., Rzeszotarski, J., & Wang, J. (2021). DataPrep.EDA: Task-Centric Exploratory Data Analysis for Statistical Modeling in Python. *Proceedings of the ACM SIGMOD International Conference on Management of Data*, 2271–2280. https://doi.org/10.1145/3448016.3457330

Perienen, A. (2020). Frameworks for ICT Integration in Mathematics Education - A Teacher's Perspective. *Eurasia Journal of Mathematics, Science and Technology Education*, *16*(6), em1845. https://doi.org/10.29333/EJMSTE/7803

Pes, B. (2020). Ensemble feature selection for high-dimensional data: a stability analysis across multiple domains. *Neural Computing and Applications*, *32*(10), 5951–5973. https://doi.org/10.1007/S00521-019-04082-3/TABLES/5

Pudjihartono, N., Fadason, T., Kempa-Liehr, A. W., & O'Sullivan, J. M. (2022). A Review of Feature Selection Methods for Machine Learning-Based Disease Risk Prediction. *Frontiers in Bioinformatics*, *2*, 927312. https://doi.org/10.3389/FBINF.2022.927312

Rajendran, S., Chamundeswari, S., & Sinha, A. A. (2022). Predicting the academic performance of middle- and high-school students using machine learning algorithms. *Social Sciences & Humanities Open*, *6*(1), 100357. https://doi.org/10.1016/J.SSAHO.2022.100357

Remeseiro, B., & Bolon-Canedo, V. (2019). A review of feature selection methods in medical applications. *Computers in Biology and Medicine*, *112*, 103375. https://doi.org/10.1016/J.COMPBIOMED.2019.103375

Salman, H., Almohsen, E., Henari, T., Shatnawi, S., Buzaboon, A., Fardan, M., & Albinali, K. (2023). Using Machine Learning and SEM to Analyze Attitudes towards adopting Metaverse in Higher Education. *2023 International Conference on Smart Applications, Communications and Networking, SmartNets 2023*. https://doi.org/10.1109/SMARTNETS58706.2023.10215936

Sambasivam, G., & Opiyo, G. D. (2021). A predictive machine learning application in agriculture: Cassava disease detection and classification with imbalanced dataset using convolutional neural networks. *Egyptian Informatics Journal*, *22*(1), 27–34. https://doi.org/10.1016/J.EIJ.2020.02.007

Saputra, J., Lawrencya, C., Saini, J. M., & Suharjito, S. (2023). Hyperparameter optimization for cardiovascular disease data-driven prognostic system. *Visual Computing for Industry, Biomedicine, and Art*, *6*(1), 1–27. https://doi.org/10.1186/S42492-023-00143-6/TABLES/1

Segura, M., Mello, J., & Hernández, A. (2022). Machine Learning Prediction of University Student Dropout: Does Preference Play a Key Role? *Mathematics 2022, Vol. 10, Page 3359*, *10*(18), 3359. https://doi.org/10.3390/MATH10183359

Selim, K. S., & Rezk, S. S. (2023). On predicting school dropouts in Egypt: A machine learning approach. *Education and Information Technologies*, *28*(7), 9235–9266. https://doi.org/10.1007/S10639-022-11571-X/FIGURES/6

Seota, S. B. W., Klein, R., & van Zyl, T. (2021). Modeling E-Behaviour, Personality and Academic Performance with Machine Learning. *Applied Sciences 2021, Vol. 11, Page 10546*, *11*(22), 10546. https://doi.org/10.3390/APP112210546

Shekar, B. H., & Dagnew, G. (2019). Grid search-based hyperparameter tuning and classification of microarray cancer data. *2019 2nd International Conference on Advanced Computational and Communication Paradigms, ICACCP 2019*.

https://doi.org/10.1109/ICACCP.2019.8882943

Spooner, A., Mohammadi, G., Sachdev, P. S., Brodaty, H., & Sowmya, A. (2023). Ensemble feature selection with data-driven thresholding for Alzheimer's disease biomarker discovery. *BMC Bioinformatics*, *24*(1). https://doi.org/10.1186/S12859-022-05132-9

Srinivas, S., & Rajendran, S. (2017). A Data-Driven Approach for Multiobjective Loan Portfolio Optimization Using Machine-Learning Algorithms and Mathematical Programming. *Big Data Analytics Using Multiple Criteria Decision-Making Models*, 175–210. https://doi.org/10.1201/9781315152653-8

Sulaiman, M. A., & Labadin, J. (2015). Feature selection based on mutual information for machine learning prediction of petroleum reservoir properties. *2015 9th International Conference on IT in Asia: Transforming Big Data into Knowledge, CITA 2015 - Proceedings*. https://doi.org/10.1109/CITA.2015.7349827

Suprapto, S. (2024). Comparative analysis of preprocessing methods for molecular descriptors in predicting anti-cathepsin activity. *South African Journal of Chemical Engineering*, *47*, 123–135. https://doi.org/10.1016/J.SAJCE.2023.11.001

Szymkowiak, A., Melović, B., Dabić, M., Jeganathan, K., & Kundi, G. S. (2021). Information technology and Gen Z: The role of teachers, the internet, and technology in the education of young people. *Technology in Society*, *65*, 101565. https://doi.org/10.1016/J.TECHSOC.2021.101565

Tang, J., Alelyani, S., & Liu, H. (2014). Feature selection for classification: A review. *Data Classification: Algorithms and Applications*, 37–64. https://doi.org/10.1201/B17320

Tharwat, A. (2018). Classification assessment methods. *Applied Computing and Informatics*, *17*(1), 168–192. https://doi.org/10.1016/J.ACI.2018.08.003/FULL/PDF

The pandas development team. (2023). *pandas-dev/pandas: Pandas*.

Torres-Vásquez, M., Hernández-Torruco, J., Hernández-Ocaña, B., & Chávez-Bosquez, O. (2021). Impact of oversampling algorithms in the classification of Guillain-Barré syndrome main subtypes. *Ingenius*, *2021*(25), 20–31. https://doi.org/10.17163/INGS.N25.2021.02

Trujillano, J., Sarria-Santamera, A., Aureli Esquerda, /, Badia, M., Palma, M., & March, J. (2008). Aproximación a la metodología basada en árboles de decisión (CART): Mortalidad hospitalaria del infarto agudo de miocardio. *Gaceta Sanitaria*, *22*(1), 65–72. https://scielo.isciii.es/scielo.php?script=sci_arttext&pid=S0213-91112008000100013&lng=es&nrm=iso&tlng=es

Tsai, C. F., & Sung, Y. T. (2020). Ensemble feature selection in high dimension, low sample size datasets: Parallel and serial combination approaches. *Knowledge-Based Systems*, *203*, 106097. https://doi.org/10.1016/J.KNOSYS.2020.106097

Turgut, Y. E., & Aslan, A. (2021). Factors affecting ICT integration in TURKISH education: a systematic review. *Education and Information Technologies*, *26*(4), 4069–4092. https://doi.org/10.1007/S10639-021-10441-2/METRICS

Villalobos-Arias, L., Quesada-López, C., Guevara-Coto, J., Martínez, A., & Jenkins, M. (2020). Evaluating hyper-parameter tuning using random search in support vector machines for software effort estimation. *PROMISE 2020 - Proceedings of the 16th ACM International Conference on Predictive Models and Data Analytics in Software Engineering, Co-Located with ESEC/FSE 2020*, 31–40. https://doi.org/10.1145/3416508.3417121

Wabang, K., Nurhayati, O. D., & Farikhin. (2022). Application of The Naïve Bayes Classifier Algorithm to Classify Community Complaints. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, *6*(5), 872–876. https://doi.org/10.29207/RESTI.V6I5.4498

Wah, Y. B., Ibrahim, N., Hamid, H. A., Abdul-Rahman, S., & Fong, S. (2018). Feature selection methods: Case of filter and wrapper approaches for maximising classification accuracy. *Volume 26, Issue 1, Pages 329 - 340*, *26*(1), 329–340.

Wang, J., Xu, J., Zhao, C., Peng, Y., & Wang, H. (2019). An ensemble feature selection method for high-dimensional data based on sort aggregation. *Systems Science & Control Engineering*, *7*(2), 32–39. https://doi.org/10.1080/21642583.2019.1620658

Weerts, H. J. P., Mueller, A. C., & Vanschoren, J. (2020). *Importance of Tuning Hyperparameters of Machine Learning Algorithms*. https://arxiv.org/abs/2007.07588v1

Yang, L., & Shami, A. (2020). On hyperparameter optimization of machine learning algorithms: Theory and practice. *Neurocomputing*, *415*, 295–316. https://doi.org/10.1016/J.NEUCOM.2020.07.061

Yoo, J. E., & Rho, M. (2020). Exploration of Predictors for Korean Teacher Job Satisfaction via a Machine Learning Technique, Group Mnet. *Frontiers in Psychology*, *11*, 441. https://doi.org/10.3389/fpsyg.2020.00441

Zhang, H., Zhang, H., Pirbhulal, S., Wu, W., & Albuquerque, V. H. C. D. (2020). Active Balancing Mechanism for Imbalanced Medical Data in Deep Learning–Based Classification Models. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, *16*(1s). https://doi.org/10.1145/3357253

Zhang, S., Zhang, C., & Yang, Q. (2003). Data preparation for data mining. *Applied Artificial Intelligence*, *17*(5–6), 375–381. https://doi.org/10.1080/713827180