This article has been accepted for publication in IEEE Transactions on Learning Technologies. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TLT.2023.3336541

Forecasting Gender in Open Education Competencies: A Machine Learning Approach

Gerardo Ibarra-Vazquez, María Soledad Ramírez-Montoya, and Mariana Buenestado-Fernández,

***Abstract*—This article aims to study the performance of machine learning models in forecasting gender based on the students' open education competency perception. Data were collected from a convenience sample of 326 students from 26 countries using the eOpen instrument. The analysis comprises 1) a study of the students' perceptions of knowledge, skills, and attitudes or values related to open education and its sub-competencies from a 30-item questionnaire using machine learning models to forecast participants' gender, 2) validation of performance through cross-validation methods, 3) statistical analysis to find significant differences between machine learning models, and 4) an analysis from explainable machine learning models to find relevant features to forecast gender. The results confirm our hypothesis that the performance of machine learning models can effectively forecast gender based on the student's perceptions of knowledge, skills, and attitudes or values related to open education competency.**

***Index Terms*—Open Education, Forecasting, Gender, Student Perception, Explainable, Machine learning, Higher Education, Educational Innovation.**

Acronyms

OER	Open Educational Resource
SDG	Sustainable Development Goal
ML	Machine Learning
DT	Decision Tree
RF	Random Forest
LGBM	Light Gradient Boosting Machine
CNN	Convolutional Neural Network
1D-CNN	One-Dimensional-Convolutional Neural Network

I. Introduction

The current international political agenda makes explicit the importance of generating processes to incorporate social justice in education, including good open education practices for the solvency of contextual education. The United Nations Educational, Scientific and Cultural Organization (UNESCO) Framework for Action gives a good account of this in recognition of its leading role in the Education 2030 Sustainable Development Agenda [1]. Sustainable Development Goal (SDG) 4 establishes international scientific communication to guarantee access to inclusive and equitable quality education and promote learning opportunities for all. Additionally, it formulates using free-access educational resources and technology without discrimination as a strategic measure. Their potential is recognized as part of the solution in the Strategy for Gender Equality in and through education [2]. More recently, UNESCO reaffirmed its commitment to these goals and strategies by helping all member states build inclusive knowledge societies through the Recommendation on Open Educational Resources (OER) [3]. The need to transfer the political orientations of international organizations as sources of innovation is frequently demanded to develop practices to improve the training of professionals and solve contextual problems related to practice in educational sciences [4].

There is an essential heterogeneity of open education practices. They have mainly focused on the production, use, and mobilization of various open educational resources [5]–[7], the analysis of the support infrastructure for teaching [8] and the organization and implementation of learning ecosystems [9]. Other practices include the performances of techniques for open learning spaces [10], innovative methodologies such as gamification [11], and Education 4.0 tools to support sustainable open education [12], [13]. In recent years, studies on the dissemination of knowledge through repositories have increased [14] and the mobilization of massive open courses on interdisciplinary topics [15] and of educational and social inclusion [16], [17]. For the development of good practices of open education, an attempt has been made to determine in the scientific literature a competency profile that allows, through processes of social and professional transfer, selecting training proposals that would positively impact the curricula, the students, the academic environment and the teaching culture [18], [19]. Open education, like other topics of educational sciences, can be studied by applying more advanced research designs and techniques. Consequently, gathering effective open education practices has made it possible to identify the competencies needed to develop them.

Artificial intelligence's rapid progress presents many opportunities within the realm of education, aligning with the strategic recommendations outlined in the Beijing Consensus on AI and education [20]. These advancements hold significant potential for enhancing educational processes through monitoring, evaluation, and research. In particular, machine learning offers a diverse range of algorithms that effectively elucidate and interpret complex data, enabling the construction of predictive models with precise outcomes tailored to the individual circumstances and progress of each participant in the study [21]. The present article makes a valuable and novel contribution to the scientific literature by examining the performance of machine learning models in forecasting gender, leveraging the students' perceptions of open education competency as a crucial factor in the analysis. By exploring

G. Ibarra-Vazquez is with the School of Engineering and Sciences, Tecnologico de Monterrey, Monterrey, NL, 64849 Mexico e-mail: gerardo.ibarra.v@tec.mx.

this intersection of machine learning and gender prediction, this research expands our understanding of how AI techniques can be applied to uncover valuable insights and facilitate informed decision-making in education.

In this research, we focus on studying the application of machine learning models to forecast gender based on students' perceptions of open education competency. We delve into these main contributions of this research. 1) The study gathered data from 326 students in 26 countries using the eOpen instrument, which provided a diverse sample, allowing for more generalizable findings and insights into variations in gender perceptions and open education competency across different contexts. 2) The research utilizes explainable machine learning models to predict gender from a 30-item questionnaire, employing cross-validation to ensure model robustness and conducting a comparative analysis of different models while enhancing interpretability by identifying relevant features contributing to gender forecasting based on open education competency perceptions. 3) Stratified k-fold cross-validation highlighted notable differences between RF-LBGM and RF-1D-CNN compared to traditional k-fold methods. 4) Machine learning models are fundamental tools for learning differences between genders and reducing open education competency gaps.

The article is organized as follows: Section II provides the background and outlines related studies on open education, the application of machine learning in education, and gender in perception studies. Section III outlines the experimental research approach. Section IV presents the findings, while Section V discusses the data in comparison to related works. Finally, Section VI concludes the article by highlighting the study's findings, limitations, implications for research and practice, and future work.

II. Related works

A. Open education

Open education is democratized through facilitators that promote the social appropriation of knowledge. According to the literature review of [22], open education is an educational model comprising the design, development, and evaluation of learning opportunities freely and with open access to improve students' quality of learning throughout life. Different frameworks have been created for the development of open education in educational institutions: the OpenEd Quality Framework [23], the Open Educators Factory Framework [24], the 6E evaluation framework [25], the Opening Up Education: A Support Framework for Higher Education Institutions [26], and the framework for selecting OER based on fitness-for-purpose [27]. At the same time, instruments have been designed for its evaluation, among which the following stand out: the Scale of Digital Competence and Use of Open Educational Resources (CD-REA) [28], the Scale of Accreditation Standards of Open and Distance Education [29], and the Scale of Perspectives and Opinions on OER and other online educational resources [30]. The connection between the institutional framework for academic research and those who transfer it to practice supposes an added value of the scientific communities who contribute to societal cultural and social development. As a transfer exercise of UNESCO's Recommendation on Open Educational Resources, the authors of [31] propose the following competency indicators:

Furthermore, it has become increasingly evident that fostering changes to advance the development of open education practices is crucial, as highlighted by the perceptions of students and education professionals regarding various socio-demographic characteristics, including gender [32]. Traditional research approaches in education have already shed light on the advancements and limitations of open education practices when examined through a gender lens. Notably, in a critical analysis conducted in [17], the texts of open education resources were scrutinized from a gender perspective, revealing the imperative to incorporate a more equitable and inclusive vision. Gender has also been a significant factor in analyzing participation rates in open education programs, motivations for engagement, time allocation to training, and students' use of computers and the internet [33]. Through these studies, a deeper understanding of the intersection between gender and open education practices has emerged, emphasizing the need for continued examination and action to foster more significant equity and inclusion in education.

B. Machine learning in education

Machine Learning (ML) is substantially changing the conventional research methods in education. ML is a subfield of Artificial Intelligence focused on developing models to enable computers to learn and make predictions without being explicitly programmed [34]. Leading international organizations such as UNESCO and Organization for Economic Cooperation and Development (OECD) have recommended the integration of artificial intelligence in educational research to improve the quality of educational systems [20], [35]. Machine learning sets itself apart from traditional linear regression through its ability to offer robust non-parametric components that enable greater adaptability and versatility in its practical applications as a predictive model [36]. These techniques allow researchers and educational experts to design more personalized curricula tailored to the needs of specific groups, optimizing teaching strategies and learning progress to improve academic performance.

Machine Learning has found diverse applications in the field of education. It has facilitated problem-solving in reasoning, knowledge representation, prediction, learning, and perception [37]. Also, these models have been used to identify patterns in massive educational data and develop predictive models [38]. This represents an advance in educational research methodology through understanding more complex phenomena than traditional approaches focused on quantitative and qualitative research methodologies that are more limited [39]. In [40], the authors point out that the benefits of machine learning include providing more accurate information about what happens in learning contexts and better knowledge of the characteristics of educational agents that make it possible to offer learning experiences. More individualized training helps teachers make better design decisions based on the needs of students.

In [41], authors point out that studies based on the application of ML in education have focused mainly on predicting

TABLE I: OER Recommendation objectives and indicators.

OER Recommendation objectives	Competency indicators according to [31].
1. Capacity building	1.1. Creation, reuse, adaptation and redistribution of OER. 1.2 Open licenses and copyright. 1.3 Digital literacy.
2. Development of supportive policies	2.1 Policies to promote open education. 2.2 Policies for privacy and data protection.
3. Effective, inclusive and equitable access to quality OER	3.1 Open access sharing programmes or technology platforms. 3.2 Development of inclusive OER. 3.3 ICT and broadband infrastructure.
4. Promoting the development of sustainability models	4.1 Sustainability models. 4.2 Funding sources and sustainability. 4.3 Linguistic translation of open licenses.
5. Promotion and facilitation of international cooperation.	5.1 Projects with international cooperation. 5.2 International funding mechanisms. 5.3 Peer networks (local, regional and global).

learning performance and school dropout rates and on online and blended learning environments of college students in Computer Science or Science, Technology, Engineering, and Mathematics (STEM). However, ML has also allowed the study of students' perceptions in the educational field. Salas-Rueda's research has made notable contributions in the areas of student's perceptions of the flipped classroom methodology, blended learning, and the use of social networks [42], [43], [43]. Additional related works have explored the application of machine learning in perception like self-regulated learning [44], the challenges and potentials of artificial intelligence in the educational process [45], the perception of emotional competence in job performance [46], and the assessment of academic quality [47]. The application of machine learning in various aspects of education, including predicting learning performance, exploring different teaching methodologies, and studying perception, has demonstrated its potential to improve educational systems and enhance students' learning experiences.

C. Gender in education

Understanding existing gender differences in skill development is becoming increasingly important to gain valuable information on how to close the gender gap. The analysis of specific sociodemographic characteristics in machine learning studies in education is still in its infancy [48], [49]. Gender prediction using machine learning models mainly focused on gender prediction of educational leadership [50]; female models and reinforcement in STEM [51]; exploring gender differences in learning computational thinking [52]; the intersection of the academic gender gap [53]; and gender stereotyping in academic dropout [54]. These studies shed light on the potential of machine learning to provide valuable insights and guide efforts toward closing the gender gap in education.

Studying students' perceptions, gender, and the integration of machine learning is essential to gaining valuable insights into educational practices and developing more inclusive and effective learning environments. Authors from [55] focused on studying students' perceptions of self-directed learning and performance in open education processes and found no significant differences in terms of gender. However, significant differences were found in choosing massive and open courses: women tended to participate in practical ICT courses and had anxiety before exams; they participated less in entrepreneurship courses [56]. Including sociodemographic variables, such as gender, in research designs provides valuable insights for developing teacher-training processes that positively impact open education practices.

Open education competency has predominantly been explored using conventional education research methods, with limited attention given to studying students' perceptions in the context of machine learning and gender. Thus, studying the perception of open education competency through machine learning models while considering sociodemographic variables like gender represents a novel and valuable contribution of machine learning to education. Advancing techniques for analyzing open education competency, particularly considering gender as a significant sociodemographic variable, is imperative to drive improvements in educational quality, foster innovation in education, broaden student choices, close gender gaps, and enhance the educational system.

III. Methodology

In this work, we propose to study the performance of machine learning models in forecasting gender based on the students' perceptions of their knowledge, skills, and attitudes or values related to open education. We used eOpen data to perform the experiments to forecast the participant's gender using four of the most widely used machine learning classification models: 1) Decision Trees (DT), 2) Light Gradient Boosting Machine (LGBM), 3) Random Forest (RF), and 4) One-Dimensional-Convolutional Neural Network (1D-CNN).

A DT is a tree-like model composed of a root node, branches, and leaf nodes. A root node represents each test of an item response, the output from the root node is a branch, and a leaf represents each class label [57]. LGBM are based on Gradient Boosting Decision Trees with an optimization method that can effectively reduce the number of features without hurting the accuracy of split-point determination [58]. Random Forest (RF) algorithms are popular machine learning algorithms in classification and regression tasks [59]. These algorithms construct a series of decision trees on different samples and use majority voting to assign a class. Convolutional Neural Network (CNN) is the most popular machine learning model for most computer vision tasks. Recently, CNNs

This article has been accepted for publication in IEEE Transactions on Learning Technologies. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TLT.2023.3336541

have been rapidly used in many one-dimension applications, becoming state-of-the-art [60].

In our experimentation, data pre-processing was considered to diminish bias acquired by machine learning models in the learning process from categories and questions with high numerical contribution [61], [62]. Then, we followed two steps: 1) oversampling the class with lower occurrences to level its size to the class with major occurrences using duplicate instances and 2) normalization, which was applied as the standard procedure over all items in eOpen data. We performed initial experimentation by dividing data into training and testing phases to evaluate the models' behavior with the most widely used metric: *accuracy*. Then, we performed two cross-validation methods to validate the initial performance of the machine learning models using the average cross-validated accuracy and statistical hypothesis test to observe differences between the machine learning models.

In machine learning, two distinct approaches influence the interpretability of models: explainable and non-explainable machine learning [63]. Explainable machine learning refers to algorithms and techniques that allow humans to comprehend the rationale behind model predictions. This approach emphasizes transparency and aims to provide explanations for the decision-making process [64]. On the other hand, non-explainable machine learning is considered a black-box model, which focuses on achieving high predictive accuracy without prioritizing interpretability [65]. These models often rely on complex architectures, such as deep neural networks, which can be challenging to interpret due to their intricate inner workings.

Three machine learning models are explainable from our experimentation due to their decision tree structure (DT, RF, and LGBM). With this understanding, we thoroughly analyzed the decision rules employed by these models. The main objective was to find the most relevant features influencing the models' ability to forecast gender accurately.

A. Cross-Validation

Cross-validation is a statistical method for evaluating and validating machine learning models using different parts of the data to train and validate the model in several operations [66]. Typically, the cross-validation process resides in subsequent stages of training and validation sets so that each instance in data has an opportunity to be validated. The most common cross-validation method is k-fold cross-validation, where data is separated into k approximately equally folds. Then, a cycle of k iterations is executed to train and validate the model using a different data fold for validation and $k-1$ folds for training. Therefore, the algorithm's behavior is observed at each fold by calculating a performance metric such as accuracy. At the end of k iterations, different methods, such as averaging or statistical tests. Figure 1 demonstrates an example of the most common k-fold cross-validation in machine learning with $k=10$. The blue-section data is used for training, while the red sections are used for validation.

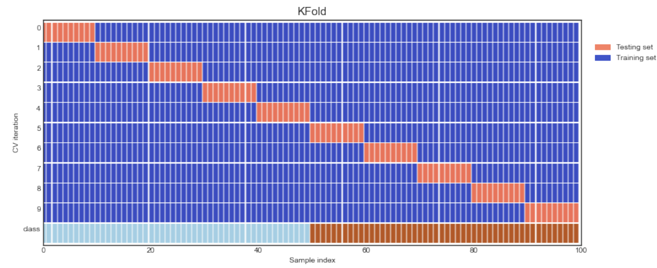

Fig. 1: Visual representation of Kfold strategy.

Nonetheless, different variations of k-fold cross-validation exist that necessitate repeated rounds of k-fold cross-validation or stratified folds; the repeated rounds ensure that each data fold has the same proportion of instances within a given label (See Figure 2). In article [67], the author recommended stratified 10-fold cross-validation as the best model selection method from a study of several approaches, where he included different cross-validation methods (regular cross-validation, leave-one-out cross-validation, and stratified cross-validation) and bootstrap to estimate the accuracy. Stratified 10-fold cross-validation yielded less tendency performance estimation.

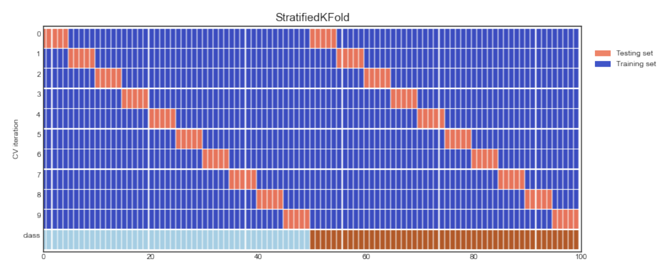

Fig. 2: Visual representation of Stratified kfold strategy.

B. Participants

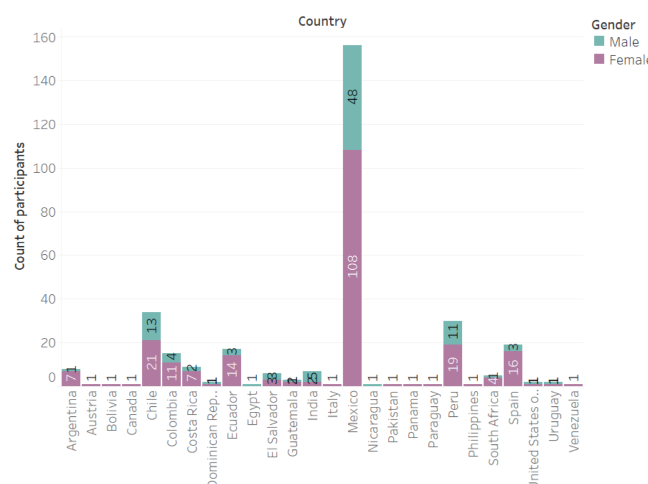

Fig. 3: Count of participants for each country. Color shows details about gender.

The data collected in [68] were from a convenience sample of $m=326$ students, of which 226 were females and 100 were males from 26 countries. The students belonged to three different academic degrees: 1) Doctoral degree, 2) Master's degree, and 3) Bachelor's degree and below. Figure 3 illustrates the geographic distribution of students, whereas Table II presents a detailed breakdown of the participants by gender and academic degree. The data collection process involved a voluntary self-assessment questionnaire administered through Google Forms with the approval of the ethics committees at the participating institutions, and informed consent was obtained from all individual participants.

TABLE II: Statistics from the participants by gender.

Academic degree	Female	Male
Doctoral degree	57	28
Master's degree	93	35
Bachelor's degree and below	76	37
Dataset	226	100

C. Instrument

The eOpen instrument is intended to be of value to academic, scientific, and social communities interested in open education, educational innovation, research evaluation, and complex environments. The instrument designed in [68] measures students' perceptions of knowledge, skills, and attitudes or values related to open education and its sub-competencies from a 30-item questionnaire:

1) Capacity building (items 1, 2, 3, 4, 5, 6, 7, 8);
2) Development of support policies (items 9, 10, 11, 12, 13, 14, 15);
3) Promotion of effective, inclusive, and equitable access (items 16, 17, 18, 19, 20, 21);
4) Creation of sustainability models (items 22, 23, 24, 25);
5) Promotion of international cooperation (items 26, 27, 28, 29, 30);

The eOpen instrument uses a four-point Likert scale (1: Strongly disagree, 2: Disagree, 3: Agree, 4: Strongly agree) to measure the students' level of self-assessment.

D. Dataset description

The eOpen dataset comprises 326 instances, encompassing 33 attributes, which include three socio-demographic variables (gender, academic degree, and country of residence), along with responses to the 30-item questionnaire. In our analysis, we utilize descriptive analysis to illustrate the participants' responses to the eOpen questionnaire. Firstly, we employ box plots to comprehensively represent the data distribution for each questionnaire item. These plots reveal key statistics such as the median (highlighted within the gray box), the interquartile range (represented by the colored box), and the presence of outliers (displayed as scattered points). This visual representation is shown in Figure 4.

We extended the data analysis beyond the scope of questionnaire responses by presenting box plots for each open education sub-competencies, as depicted in Figure 5. These box plots were constructed by averaging the responses to each sub-competency across all participants, illustrating the distinct behavioral patterns associated with each sub-competency.

Lastly, we conducted a gender-based analysis of the eOpen dataset, employing fundamental descriptive statistics for each open education sub-competency. This analysis involved computing both the mean and standard deviation for each sub-competency. Table III presents the variations in sub-competency scores between genders. Notably, we observed more significant variance in scores among females than males despite the nearly identical mean values.

TABLE III: Descriptive statistics from eOpen data by sub-competency.

Gender		Capacity	Development	Promotion	Creation	Intl. Cooperation
Female	Mean	2.93	2.90	2.97	2.68	2.70
	Std.	0.49	0.49	0.46	0.62	0.54
Male	Mean	2.97	2.98	3.01	2.71	2.78
	Std.	0.62	0.60	0.61	0.71	0.65

IV. RESULTS

Table IV shows the dataset through the pre-processing procedure (data normalization and oversampling) and data split. Firstly, we used the standard data normalization and randomly oversampled the minority class (male class) to level the classes in the dataset by adding random duplicated instances from this class. The data ended with 220 samples in each class, as shown in Table IV. Then, we split the data to produce the training and testing datasets for the classification algorithms. The *training* and *testing* columns show the instances for each class in each dataset, approximately 90% and 10%, respectively.

TABLE IV: Dataset configuration.

Gender	Original	Balanced data	Training	Testing
Female	226	226	209	17
Male	100	226	197	29

For the machine learning models' implementation, we use scikit-learn v0.21.2 for DT and RF, lightgbm v3.3.5 for LGBM, and keras v2.3.1 for a custom-made 1D-CNN. Table V shows the layer configuration for 1D-CNN. Therefore, we enlisted next the empirical hyperparameters' tune for training each model:

- **DT**: default setting with Gini impurity criterion for the Shannon information gain.
- **LGBM**: number of estimators=*100*.
- **RF**: number of trees in the forest=*1000*, maximum depth of the tree=*15*.
- **1D-CNN**: loss function=*sparse categorical crossentropy*, optimizer=*Adamax*, metric=*sparse categorical accuracy*, batchsize=*5*, epochs=*20*.

TABLE V: 1D-CNN layers' configuration.

1D-CNN layers
input (25×1)
1D-Convolutional layer of 64 filters, and kernel size = 5
Maxpooling layer of pool size = 5
Fully Connected layer of 1000 neurons
Fully Connected layer of 2 neurons
soft-max layer

Table VI shows the corresponding computed accuracy for each model at each training and testing stage with the eOpen

This article has been accepted for publication in IEEE Transactions on Learning Technologies. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TLT.2023.3336541

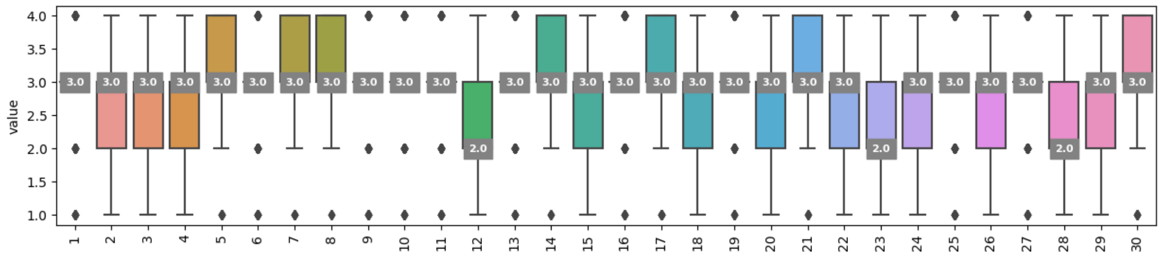

Fig. 4: Box Plots from the 30-Item eOpen Questionnaire. Median values are shown in the gray box.

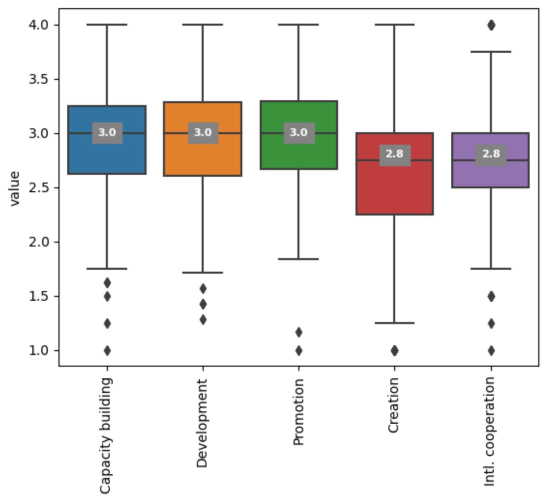

Fig. 5: Box Plots from the average of each sub-competency from eOpen Questionnaire. Median values are shown in the gray box.

data. In this Table, We observed almost similar behavior among all the machine learning models in the training phase; they scored around 95% accuracy. In the testing stage, the performance of all classifiers behaved differently. The most accurate model was RF, with 86.96% correct gender predictions. Then, DT attained 84.78% accuracy, 1D-CNN attained 71.73%, and LGBM attained 65.22%.

Our results showed differences in two optimized alternatives to DT, which obtained contrasting outcomes. On the one hand, RF's architecture improved accuracy by 2% creating a set of DTs to enable a voting decision about the participants' gender. On the other hand, the optimization made by LGBM, which reduced the number of features of split-point determination, decreased accuracy by approximately 19%. Differently, 1D-CNN was below the mean accuracy of all the test results (77.17%).

TABLE VI: Results from computing the rate of correct predictions (*accuracy*) of all models in the training and testing stages.

Algorithm	Training accuracy	Testing Accuracy
DT	97.78%	84.78%
LGBM	94.58%	65.22%
RF	97.78%	86.96%
1D-CNN	95.32%	71.73%

Then, we performed traditional and stratified k-fold cross-validation methods to estimate the performance of the machine learning models using the available data to indicate generalizability. This refers to the model's ability to adapt perfectly to new data that has not been previously used in the learning and testing stages with the same distribution as the data used to build the model. Table VII shows the computed accuracy in each round of the traditional k-fold cross-validation with $k = 10$. The last two columns exhibit the mean and standard deviation from the ten training and validation scores rounds.

Our results demonstrated a change between the first results and the application of the traditional k-fold cross-validation method. First, LGBM and DT abruptly change their respective scores. The LGBM's mean accuracy was approximately 11% superior to its first record from Table VI. DT performance was below 6% from its first result to the mean accuracy. Differently, the RF and 1D-CNN mean accuracies remained almost equal to their first score, only with a shift below 3.56%. In terms of standard deviation, models presented different behaviors. DT and 1D-CNN obtained the highest variances in the cross-validation method with approximately 13%, meaning such models have less generalizable performance. RF exhibited the lowest variance between the models with 8.6%, and LGBM attained 9.8%.

TABLE VII: Results from computing the rate of correct predictions (*accuracy*) of all models in each phase of the traditional k-fold cross-validation. The last columns denote the mean accuracy and standard deviation from the experimentation data.

Fold	1	2	3	4	5	6	7	8	9	10	Mean	Std.
DT	84.78	63.04	71.11	71.11	71.11	68.89	66.67	93.33	97.78	100	78.80	13.10
LGBM	71.74	73.91	68.89	75.56	71.11	68.89	64.44	84.44	95.56	91.11	76.60	9.80
RF	78.26	76.09	80.00	82.22	75.56	75.56	80.00	88.89	97.78	100	83.40	8.60
1D-CNN	67.39	54.34	80.00	68.89	68.89	57.77	84.44	77.77	95.55	93.33	74.80	13.20

Afterward, we experimented using a stratified k-fold cross-validation method with $k = 10$. Table 2 shows the computed scores in each round of the cross-validation method. The last two columns exhibit the mean and standard deviation from the ten training and validation scores rounds. These results display to traditional k-fold cross-validation. However, all machine learning models showed different behavior regarding standard deviation.

All models presented a stable performance with this cross-validation method, obtaining approximately 5.4% of their performance variance. Therefore, mean accuracy will reveal a better understanding of the behavior of the machine learning models. DT, LGBM, and 1D-CNN attained 74.8-78.1% compared to Rf, which presented 83.9% mean accuracy. RF exhibited a stable performance in both cross-validation methods and did not change highly from the first experimentation illustrated in Table VI. The rest of the models presented significant changes from 4.6% up to 9.6% from the first experimentation.

TABLE VIII: Results from computing the rate of correct predictions (*accuracy*) of all models in each phase of the stratified *k*-fold cross-validation. The last columns denote the mean accuracy and standard deviation from the experimentation data.

Fold	1	2	3	4	5	6	7	8	9	10	Mean	Std.
DT	69.57	71.74	73.33	77.77	80.00	82.22	82.22	86.66	75.56	82.22	78.10	5.20
LGBM	67.39	65.22	75.56	73.33	73.33	75.56	75.56	84.44	75.56	82.22	74.80	5.50
RF	73.91	76.09	82.22	84.44	82.22	84.44	86.66	88.89	88.89	91.11	83.90	5.30
1D-CNN	65.22	76.08	68.88	73.33	82.22	77.77	77.77	80.00	80.00	82.22	76.40	5.40

Using the cross-validation methods to find significant differences between all the machine learning models, we first performed a statistical test to verify the data homogeneity assumption using Levene's test. We obtained for the traditional *k*-fold cross-validation method: $F(3,36) = 1.704,\ p = 0.183$. For the stratified *k*-fold cross-validation method, the outcome was: $F(3,36) = 0.043,\ p = 0.980$. Homogeneity of variances was admitted in the two methods, which allowed the ANOVA test to be applied. Table IX shows the results after performing the ANOVA test in both cross-validation methods data. Data were obtained for the stratified *k*-fold cross-validation method; a *p*-value was below 0.05, indicating significant differences between the machine learning models using such a method. The traditional *k*-fold method showed a *p*-value above 0.05, which denoted no significant differences between the models.

TABLE IX: Results after applying the ANOVA test. Mean and Std. denotes the mean and standard deviation. F denotes a value on the F distribution. Df presents the degrees of freedom. *p* is used to quantify the statistical significance of a result.

Cross-validation method	ML Model	Mean	Std.	F	df	*p*
Traditional *k*-fold	DT	.78	.138	.97	3	.418
	LGBM	.76	.103			
	RF	.83	.090			
	1D-CNN	.74	.138			
Stratified *k*-fold	DT	.78	.054	4.97	3	.005
	LGBM	.74	.057			
	RF	.83	.056			
	1D-CNN	.76	.056			

As the stratified *k*-fold cross-validation method data presented significant differences between machine learning models, the Bonferroni post hoc test was applied to identify the pairs of means in which statistical significance was achieved. Table X presents the Bonferroni post hoc test results. We observed that LGBM-RF and RF-1D-CNN were significantly different between each pair of group means.

TABLE X: Results after applying the Bonferroni post hoc test. The mean difference denote the subtraction from each pair of group means. *p* is used to quantify the statistical significance of a result.

Comparison		Mean difference	*p*
DT	LGBM	.033	1.000
	RF	.057	.168
	1D-CNN	.017	1.000
LGBM	RF	.090	.005
	1D-CNN	.015	1.000
RF	1D-CNN	.075	.029

Next, we will present the analysis results using explainable machine learning models to identify the relevant characteristics for forecasting gender based on open education competencies. We will discuss the top features derived from the three decision tree models, and for the complete decision trees, please refer to the Appendix section. Figure 6 presents the DT model's top features that contribute the most to gender prediction based on the student's open education competencies. The most critical component in the model's decision rule is $X[28]$, corresponding to question 29 regarding the *promotion of international cooperation* subcompetency. In the second level in the hierarchy, we found questions 25 and 4, which refer to *creation of sustainability models* and *capacity of building* sub-competencies.

Figure 7 illustrates the top features from the decision rules of the LGBM model. The most important feature for this model to predict gender was *Column*_28 that represents question 29 relating to the *promotion of international cooperation* subcompetency. Furthermore, the hierarchical structure analysis revealed additional relevant features, including questions 6 and 1, which correspond to the *capacity of building* subcompetency.

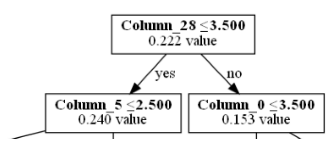

Fig. 7: Top features from tree model generated from Light Gradient Boosting Machine model.

Figure 8 exhibits the top-level RF decision rules, highlighting the most pertinent features according to this model. The decision rule's most critical feature is $X[27]$, representing question 28 associated with the *promotion of international cooperation* subcompetency. Continuing the analysis in the hierarchical structure, we discovered questions 17 and 3, which align with the sub-competencies of *promotion of effective, inclusive, and equitable access* and *capacity of building*, respectively.

V. Discussion

The study of open education competency data through machine learning models expands data analysis capabilities to extract characteristics from participants' perceptions to predict gender. In Table VI, we observed that machine learning models could predict the participants' gender from the

This article has been accepted for publication in IEEE Transactions on Learning Technologies. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TLT.2023.3336541

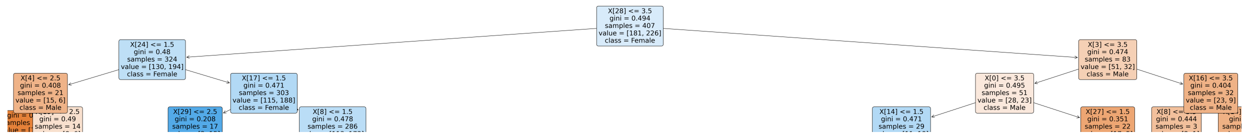

Fig. 6: Top features from tree model generated from Decision Tree model.

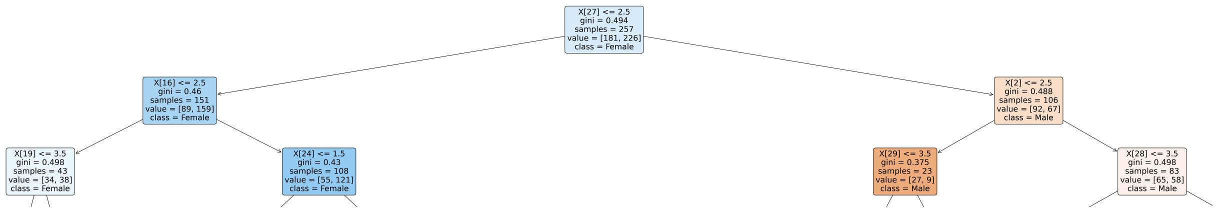

Fig. 8: Top features from a tree model generated from Random Forest model.

student's perception of knowledge, skills, and attitudes or values related to open education using the eOpen instrument. This finding agrees with [69], which employed a discriminant characteristic to predict student performance using machine learning algorithms. Thus, machine learning models enhanced data analysis from survey researchers by extracting features from students' perceptions about open education to forecast the participants' gender.

Cross-validation methods presented more accurate performance estimation for the machine learning algorithms and revealed which algorithms were more likely to generalize the models. In Tables VII and VIII, we observed that both methods provide a better understanding of the models' behavior than the initial experimentation. The mean and standard deviation demonstrated which methods have a stable performance. Authors from [66] recommended cross-validation methods to measure the generalizability of machine learning models. Cross-validation techniques improve the model selection for predicting gender using open education competency data.

Stratified k-fold cross-validation was demonstrated to reveal significant differences between RF-LBGM and RF-1D-CNN with the multiple comparisons of group means using statistical tests compared to traditional k-fold cross-validation. Table IX shows significant differences for stratified k-fold cross-validation. This argument supports the research article from [67], which recommended stratified 10-fold cross-validation as the best model selection method in a study of several validation approaches.

Statistical analysis found significant differences between the performances of machine learning models. In Table X, we demonstrated the discovery of significant differences between RF-LGBM and RF-1D-CNN through statistical tests. Authors such as [70], [71] also proposed using k-fold cross-validation followed by an appropriate hypothesis test rather than directly comparing the average accuracy to compare different machine learning models. Therefore, the comparison between machine learning models using cross-validation methods can be complemented using proper statistical tests instead of directly comparing the average accuracy.

Explainable machine learning models like DT, RF, and LGBM employ decision rules to determine the most important features to predict the student's gender based on their perception of open education competency. The most relevant features used by these models are visually presented in Figures 6-8, providing a comprehensive overview of the explainable machine learning models' findings. It is noteworthy that all of these models converge on the *promotion of international cooperation* subcompetency as a crucial factor in gender prediction. This convergence highlights the consistent importance of this particular subcompetency across the various models and underscores its significant influence on gender prediction. Understanding these significant features can provide valuable insights into students' perceptions, enabling the development of improved educational policies and strategies. Compared to a non-explainable machine learning model like 1D-CNN, Random Forest (RF) demonstrates superior accuracy in gender prediction, making it a reliable tool for understanding students' open education competency perceptions and their connection to gender.

VI. Conclusion

Over the past few years, an increasing focus has been on open education to make learning accessible to all and foster equality among learners. Effective measuring and evaluating open education competencies remains challenging despite its potential benefits. This article aims to study the performance of machine learning models in forecasting gender based on the students' open education competency perception. Our findings indicate that machine learning models can effectively predict participants' gender from students' perceptions of knowledge, skills, and attitudes or values related to open education and its sub-competencies. This discovery holds great significance as it demonstrates the potential of machine learning to offer valuable insights into learners' competencies perceptions, thereby informing educational practices.

In this study, we considered the following data analyses: 1) study the students' perceptions of knowledge, skills, and attitudes or values related to open education and its sub-competencies with a 30-item questionnaire, using machine learning models to forecast participants' gender, 2) validation performance through cross-validation methods, 3) statistical analysis to find significant differences between machine learning models, and 4) an analysis from decision trees to find relevant features to forecast gender. Machine learning demonstrated the capability to build models that can predict the gender of unknown participants' perceptions of open education with up to 84% mean accuracy. Therefore, the mathematical models built with machine learning extracted features from eOpen data and found differences in the perceptions by gender. Additionally, the higher accuracy exhibited by RF in gender prediction highlighted its reliability as a valuable tool for understanding students' open education competency perceptions and their relationship to gender. Using explainable machine learning models, such as DT, RF, and LGBM, allowed to identify significant features in predicting students' gender based on their perceptions of open education competency, providing valuable insights for educational strategies. Additionally, we could contrast functionalities and performance between explainable and non-explainable models.

While this study effectively showcases the potential of machine learning models in analyzing perception data and predicting competency levels, it is essential to acknowledge the study's limitations. Even though the study involved a series of cross-validation methods in verifying the machine learning models' performance, it is crucial to recognize that further research is needed to establish the generalizability of the findings with larger populations. Another limitation is that decision trees only provide insights into the features selected by the model for constructing decision rules, which necessitates interpretation and explanation by researchers.

This work has practical and research implications, demonstrating the utility of machine learning models in predicting participants' gender based on students' perceptions of their knowledge, skills, and attitudes/values related to open education. This has practical value for educators, enabling them to identify areas where students may require additional support or resources. The study also emphasizes the significance of utilizing perception data and machine learning techniques' efficiency in accurately analyzing this data. Our contributions expand the body of literature on machine learning applications in education. Our findings indicate that machine learning models can effectively analyze perception data to make accurate predictions regarding gender. Additionally, the study underscores the importance of comprehending the decision-making process of explainable machine learning models. This work establishes a foundation for future research endeavors exploring the utilization of machine learning in education perception data and emphasizes its potential for advancing our understanding of students' competencies.

Appendix

Acknowledgment

References

[1] United Nations Educational, Scientific and Cultural Organization, *Educación 2030: Declaración de Incheon y Marco de Acción para la realización del Objetivo de Desarrollo Sostenible 4: Garantizar une aducación inclusiva y equitativa de calidad y promover oportunidades de aprendizaje permanente para todos*. UNESCO Publishing, 2017. [Online]. Available: https://unesdoc.unesco.org/ark:/48223/pf0000245656_spa

[2] ——, *Del accesso al empoderamiento: estrategia de la UNESCO para la igualdad de género en y a través de la educación 2019-2025*. UNESCO Publishing, 2019. [Online]. Available: https://unesdoc.unesco.org/ark:/48223/pf0000371127

[3] ——, *Proyecto de recomendación sobre los Recursos Educativos Abiertos (REA)*. UNESCO Publishing, 2019. [Online]. Available: https://unesdoc.unesco.org/ark:/48223/pf0000370936_spa

[4] O. Fadiran, J. van Biljon, and M. Schoeman, "Knowledge transfer in science education: The case for usability-based knowledge visualization guidelines," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 12067 LNCS, pp. 263–273, 2020.

[5] G. Cox, M. Willmers, and B. Masuku, "Sustainable open textbook models for social justice," *Frontiers in Education*, vol. 7, 2022.

[6] A. Kaur, S. Gulati, R. Sharma, A. Sinhababu, and R. Chakravarty, "Visual citation navigation of open education resources using litmaps," *Library Hi Tech News*, vol. 39, no. 5, pp. 7–11, 2022.

[7] A. Kilpatrick, F. Rahman, A. Anjum, S. Shome, K. Andalib, S. Banik, S. Chowdhury, P. Coombe, Y. Astroz, J. Douglas, P. Eranti, A. Kiran, S. Kumar, H. Lim, V. Lorenzi, T. Lubiana, S. Mahmud, R. Puche, A. Rybarczyk, S. Al Sium, D. Twesigomwe, T. Zok, C. Orengo, I. Friedberg, J. Kelso, and L. Welch, "Characterizing domain-specific open educational resources by linking iscb communities of special interest to wikipedia," *Bioinformatics*, vol. 38, pp. I19–I27, 2022.

[8] A. Risquez, C. McAvinia, Y. Desmond, C. Bruen, D. Ryan, and A. Coughlan, "Towards a devolved model of management of oer? the case of the irish higher education sector," *International Review of Research in Open and Distance Learning*, vol. 21, no. 1, pp. 82–98, 2020.

[9] D. Otto and M. Kerres, "Increasing sustainability in open learning: Prospects of a distributed learning ecosystem for open educational resources," *Frontiers in Education*, vol. 7, 2022.

[10] C. Ochoa-Arnedo, A. Flix-Valle, A. Casellas-Grau, N. Casanovas-Aljaro, O. Herrero, E. Sumalla, M. de Frutos, A. Sirgo, A. Rodríguez, G. Campos, Y. Valverde, N. Travier, and J. Medina, "An exploratory study in breast cancer of factors involved in the use and communication with health professionals of internet information," *Supportive Care in Cancer*, vol. 28, no. 10, pp. 4989–4996, 2020.

[11] N. Padilla-Zea, D. Burgos, A. García-Holgado, F. García-Peñalvo, M. Harquevaux, C. de-la Higuera, J. Brunton, and A. Tlili, "Catch the open! a gamified interactive immersion into open educational practices for higher education educators," *Frontiers in Psychology*, vol. 13, 2022.

[12] C. Chaka, "Is education 4.0 a sufficient innovative, and disruptive educational trend to promote sustainable open education for higher education institutions? a review of literature trends," *Frontiers in Education*, vol. 7, 2022.

[13] L. González-Pérez, M. Ramírez-Montoya, and F. García-Peñalvo, "Technological enablers 4.0 to drive open science and education: Input to unesco recommendations [habilitadores tecnológicos 4.0 para impulsar la educación abierta: aportaciones para las recomendaciones de la unesco]," *RIED-Revista Iberoamericana de Educacion a Distancia*, vol. 25, no. 2, pp. 23–48, 2022.

[14] X. Wang, T. Chen, Y. Zhang, and H. Yang, "Implications of the delphi method in the evaluation of sustainability open education resource repositories," *Education and Information Technologies*, vol. 26, no. 4, pp. 3825–3844, 2021.

[15] S. Martínez-Pérez, J. Cabero-Almenara, J. Barroso-Osuna, and A. Palacios-Rodríguez, "T-mooc for initial teacher training in digital competences: Technology and educational innovation," *Frontiers in Education*, vol. 7, 2022.

[16] P. Arinto, C. Hodgkinson-Williams, and H. Trotter, *OER and OEP in the global south: Implications and recommendations for social inclusion*, 2017.

[17] J. Lapum, A. Bailey, O. St-Amant, J. Garmaise-Yee, M. Hughes, and S. Mistry, "Equity, diversity, and inclusion in open educational resources: An interpretive description of students' perspectives," *Nurse Education Today*, vol. 116, 2022.

[18] A. Berland, K. Capone, L. Etcher, H. Ewing, S. Keating, and M. Chickering, "Open education resources to support the who nurse educator core competencies," *International Nursing Review*, vol. 67, no. 2, pp. 282–287, 2020.

[19] S. Tudor, "The open resurces and their influences on the formation of specific competencies for the teaching profession," 2019.

[20] United Nations Educational, Scientific and Cultural Organization, *Beijing Consensus on Artificial Intelligence and Education*. UNESCO Publishing, 2019. [Online]. Available: https://unesdoc.unesco.org/ark:/48223/pf0000368303

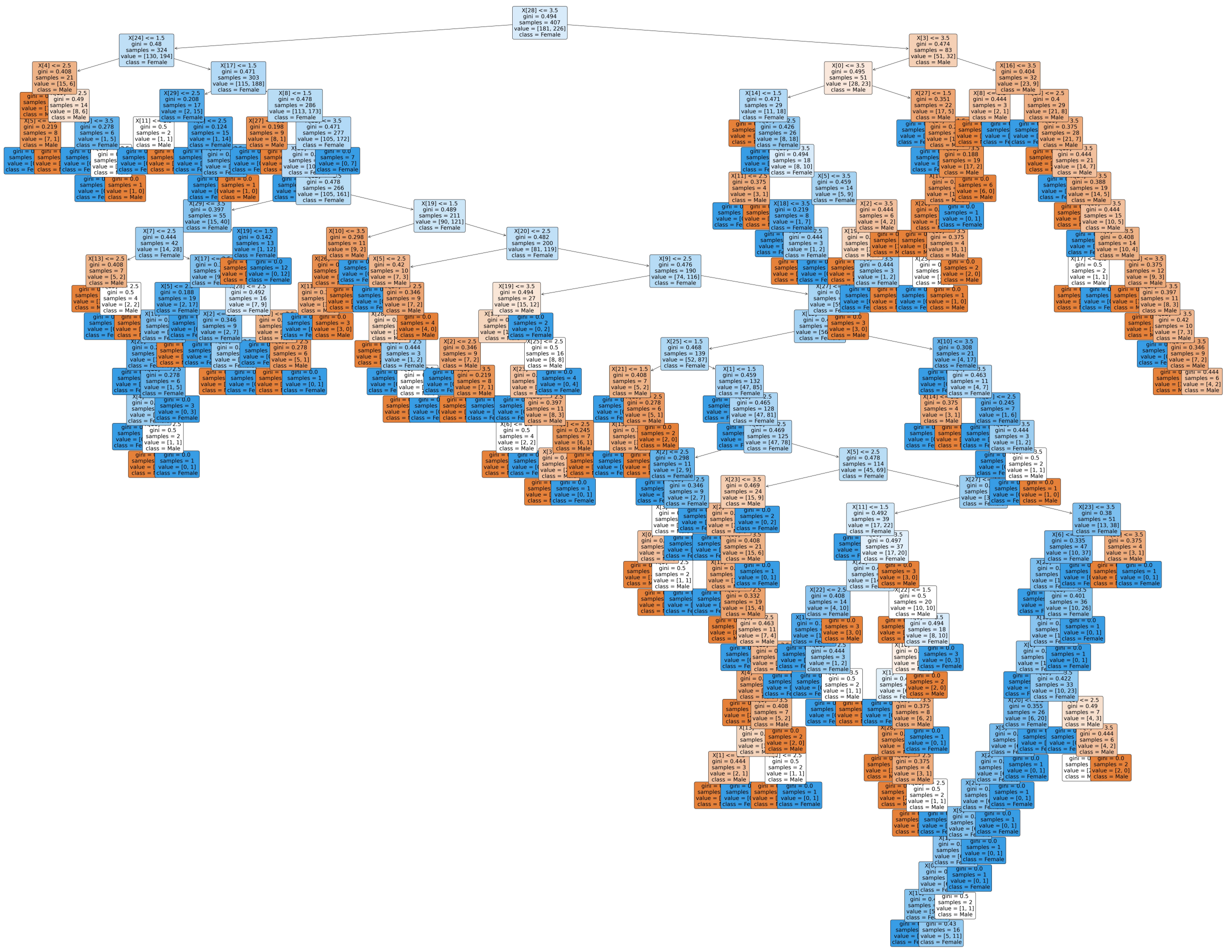

Fig. 9: DT tree model generated from eOpen data.

[21] X. Zhai, "Practices and theories: How can machine learning assist in innovative assessment practices in science education," *Journal of Science Education and Technology*, vol. 30, no. 2, pp. 139–149, 2021.

[22] M. Ramirez-Montoya, "Challenges for open education with educational innovation: A systematic literature review," *Sustainability (Switzerland)*, vol. 12, no. 17, 2020.

[23] C. Stracke, "Quality frameworks and learning design for open education," *International Review of Research in Open and Distance Learning*, vol. 20, no. 2, pp. 180–203, 2019.

[24] F. Nascimbeni and D. Burgos, "In search for the open educator: Proposal of a definition and a framework to increase openness adoption among university educators," *International Review of Research in Open and Distance Learning*, vol. 17, no. 6, pp. 1–17, 2016.

[25] Y. Geng and N. Zhao, "Measurement of sustainable higher education development: Evidence from china," *PLoS ONE*, vol. 15, no. 6, 2020.

[26] A. I. dos Santos, Y. Punie, J. C. Muñoz *et al.*, "Opening up education: A support framework for higher education institutions," Joint Research Centre (Seville site), Tech. Rep., 2016.

[27] I. Jung, T. Sasaki, and C. Latchem, "A framework for assessing fitness for purpose in open educational resources," *International Journal of Educational Technology in Higher Education*, vol. 13, no. 1, 2016.

[28] C. Sarango-Lapo, J. Mena, M.-S. Ramírez-Montoya, and E. Real, "The scale of digital competence and utilization of open educational resources (cd-rea): Factors associated with competence in bimodal university teachers [la escala de competencia digital y uso de recursos educativos abiertos (cd-rea): factores asociados a la competencia de los docentes universitarios bimodales]," *RISTI - Revista Iberica de Sistemas e Tecnologias de Informacao*, vol. 2020, no. E28, pp. 545–558, 2020.

[29] E. Can, "Open and distance education acreditation standards scale: Validity and reliability studies," *International Journal of Environmental and Science Education*, vol. 11, no. 14, pp. 6344–6356, 2016.

[30] A. Pounds and J. Bostock, "Open educational resources (oer) in higher education courses in aquaculture and fisheries: opportunities, barriers, and future perspectives," *Aquaculture International*, vol. 27, no. 3, pp. 695–710, 2019.

[31] M.-S. Ramirez-Montoya and G.-C. Tenorio-Sepulveda, "Expert validation of e-open for assessing open education competencies: An instrument designed within the framework of unesco recommendations," 2021, pp. 629–633.

[32] I. Tarling and S. Gunness, "Educators' beliefs, perceptions and practices around self-directed learning, assessment and open education practices," *Lecture Notes in Educational Technology*, pp. 187–209, 2021.

[33] M. Korkmaz, A. Özyürek, and A. Yücel, "An examination of the students enrolled in the faculties of economics, business administration and open education of anadolu university from a gender perspective," *Advances in Environmental Biology*, vol. 9, no. 14, pp. 150–161, 2015.

[34] C. M. Bishop, *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Berlin, Heidelberg: Springer-Verlag, 2006.

[35] Z. Zhongming, L. Linong, Y. Xiaona, Z. Wangqiang, L. Wei *et al.*, "Oecd digital education outlook 2021," 2021.

[36] A. D. Kumar, R. P. Selvam, and K. S. Kumar, "Review on prediction algorithms in educational data mining," *International Journal of Pure and Applied Mathematics*, vol. 118, no. 8, pp. 531–537, 2018.

[37] S. Manikandan and M. Chinnadurai, "Evaluation of students' performance in educational sciences and prediction of future development using tensorflow," *International Journal of Engineering Education*, vol. 36, no. 6, pp. 1783–1790, 2020.

[38] C. Korkmaz and A.-P. Correia, "A review of research on machine learning in educational technology," *Educational Media International*, vol. 56, no. 3, pp. 250–267, 2019.

[39] D. Gibson and D. Ifenthaler, *Preparing the next generation of education researchers for big data in higher education*, 2016.

[40] H. Munir, B. Vogel, and A. Jacobsson, "Artificial intelligence and ma-

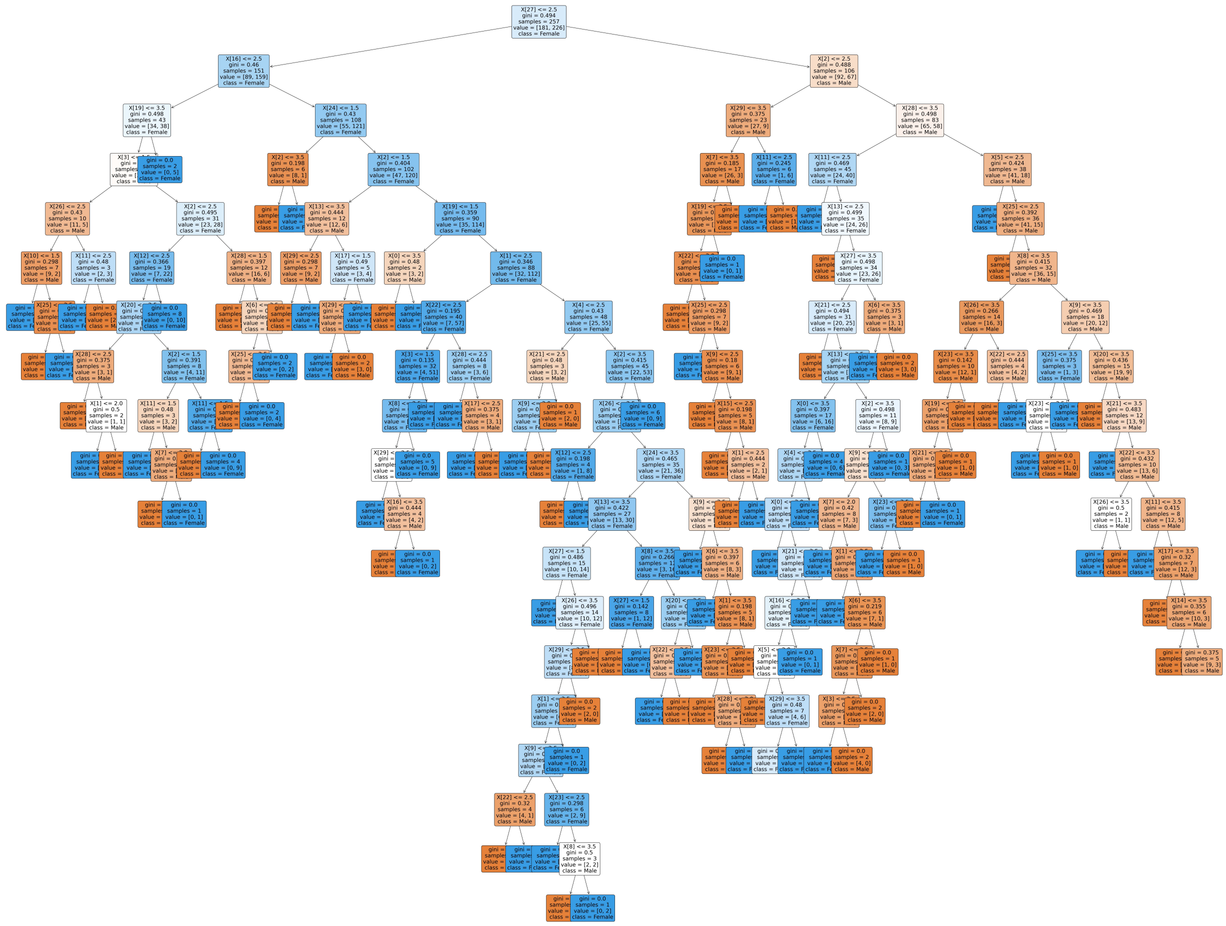

Fig. 11: RF tree model generated from eOpen data.

chine learning approaches in digital education: A systematic revision," *Information (Switzerland)*, vol. 13, no. 4, 2022.

[41] H. Luan and C.-C. Tsai, "A review of using machine learning approaches for precision education," *Educational Technology and Society*, vol. 24, no. 1, pp. 250–266, 2021.

[42] R. Salas-Rueda, "Impact of flipped classroom in the teaching-learning process on karnaugh's maps [impacto da sala de aula invertida no processo de ensino-aprendizagem nos mapas de karnaugh]," *Revista Electronica Educare*, vol. 25, no. 2, pp. 1–22, 2021.

[43] R.-A. Salas-Rueda, "Perceptions of students about the use of facebook and twitter in the educational context through data science and machine learning [percepciones de los estudiantes sobre el uso de facebook y twitter en el contexto educativo por medio de la ciencia de datos y el aprendizaje automático]," *Pixel-Bit, Revista de Medios y Educacion*, vol. 58, pp. 91–115, 2020.

[44] W. Nuankaew, P. Nuankaew, D. Teeraputon, K. Phanniphong, and S. Bussaman, "Perception and attitude toward self-regulated learning of thailand's students in educational data mining perspective," *International Journal of Emerging Technologies in Learning*, vol. 14, no. 9, pp. 34–49, 2019.

[45] R. Bucea-Manea-ţoniş, V. Kuleto, S. Gudei, C. Lianu, C. Lianu, M. Ilić, and D. Păun, "Artificial intelligence potential in higher education institutions enhanced learning environment in romania and serbia," *Sustainability (Switzerland)*, vol. 14, no. 10, 2022.

[46] R. Rasiah, J. Turner, and Y. Ho, "The impact of emotional intelligence on work performance: Perceptions and reflections from academics in malaysian higher educationobitat endiaest que," *Contemporary Economics*, vol. 13, no. 3, pp. 269–282, 2019.

[47] A. G. Perales, F. Liébana-Cabanillas, J. Sánchez-Fernández, and L. J. Herrera, "Assessing university students' perception of academic quality using machine learning," *Applied Computing and Informatics*, 2020.

[48] T. Wotaifi and E. Al-Shamery, "Mining of completion rate of higher education based on fuzzy feature selection model and machine learning techniques," *International Journal of Recent Technology and Engineering*, vol. 8, no. 2 Special Issue 10, pp. 393–400, 2019.

[49] S. Shilbayeh and A. Abonamah, "Predicting student enrolments and attrition patterns in higher educational institutions using machine learning," *International Arab Journal of Information Technology*, vol. 18, no. 4, pp. 562–567, 2021.

[50] G. Bathla, H. Aggarwal, and R. Rani, "Evolution of data analytics techniques: From data mining to big data mining," *International Journal of Advanced Science and Technology*, vol. 28, no. 19, pp. 779–795, 2019.

[51] S. Torres-Ramos, N. Fajardo-Robledo, L. Pérez-Carrillo, C. Castillo-Cruz, P. Retamoza-Vega, V. Rodríguez-Betancourtt, and C. Neri-Cortés, "Mentors as female role models in stem disciplines and their benefits," *Sustainability (Switzerland)*, vol. 13, no. 23, 2021.

[52] H. Gao, L. Hasenbein, E. Bozkir, R. Göllner, and E. Kasneci, "Exploring gender differences in computational thinking learning in a vr classroom: Developing machine learning models using eye-tracking data and explaining the models," *International Journal of Artificial Intelligence in Education*, pp. 1–26, 2022.

[53] J. Mittleman, "Intersecting the academic gender gap: The education of lesbian, gay, and bisexual america," *American Sociological Review*, vol. 87, no. 2, pp. 303–335, 2022.

[54] S. Andrade-Zurita, S. Armas-Arias, R. Núñez-López, and J. Arévalo-Peralta, "Decision trees and gender stereotypes in university academic desertion," in *Third International Conference on Image Processing and Capsule Networks: ICIPCN 2022*. Springer, 2022, pp. 321–331.

[55] V. Gooria, P. Appavoo, U. Bhunjun, and A. Gokhool, "Self-directed learning: Readiness of secondary school students in mauritius," *Lecture Notes in Educational Technology*, pp. 251–275, 2021.

[56] J. Castaño-Muñoz, K. Kreijns, M. Kalz, and Y. Punie, "Does digital competence and occupational setting influence mooc participation? evidence from a cross-course survey," *Journal of Computing in Higher Education*, vol. 29, no. 1, pp. 28–46, 2017.

[57] H. Sharma and S. Kumar, "A survey on decision tree algorithms of

This article has been accepted for publication in IEEE Transactions on Learning Technologies. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TLT.2023.3336541

classification in data mining," *International Journal of Science and Research (IJSR)*, vol. 5, no. 4, pp. 2094–2097, 2016.

[58] G. Ke, Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, Q. Ye, and T.-Y. Liu, "Lightgbm: A highly efficient gradient boosting decision tree," *Advances in neural information processing systems*, vol. 30, 2017.

[59] L. Breiman, "Random forests," *Machine Learning*, vol. 45, pp. 5–32, 2001.

[60] S. Kiranyaz, O. Avci, O. Abdeljaber, T. Ince, M. Gabbouj, and D. J. Inman, "1d convolutional neural networks and applications: A survey," *Mechanical systems and signal processing*, vol. 151, p. 107398, 2021.

[61] P. Kumar, R. Bhatnagar, K. Gaur, and A. Bhatnagar, "Classification of imbalanced data: review of methods and applications," in *IOP Conference Series: Materials Science and Engineering*, vol. 1099, no. 1. IOP Publishing, 2021, p. 012077.

[62] D. Singh and B. Singh, "Investigating the impact of data normalization on classification performance," *Applied Soft Computing*, vol. 97, p. 105524, 2020.

[63] V. Belle and I. Papantonis, "Principles and practice of explainable machine learning," *Frontiers in big Data*, p. 39, 2021.

[64] N. Burkart and M. F. Huber, "A survey on the explainability of supervised machine learning," *Journal of Artificial Intelligence Research*, vol. 70, pp. 245–317, 2021.

[65] L. Wells and T. Bednarz, "Explainable ai and reinforcement learning—a systematic review of current approaches and trends," *Frontiers in artificial intelligence*, vol. 4, p. 550030, 2021.

[66] P. Refaeilzadeh, L. Tang, and H. Liu, "Cross-validation." *Encyclopedia of database systems*, vol. 5, pp. 532–538, 2009.

[67] R. Kohavi, "A study of cross-validation and bootstrap for accuracy estimation and model selection," in *Ijcai*, vol. 14, no. 2. Montreal, Canada, 1995, pp. 1137–1145.

[68] M. S. Ramírez-Montoya and G. C. Tenorio-Sepulveda, "e-open assessment of open education competences: Validation of instrument for complex and innovative environments," 2022.

[69] Y. Altujjar, W. Altamimi, I. Al-Turaiki, and M. Al-Razgan, "Predicting critical courses affecting students performance: a case study," *Procedia Computer Science*, vol. 82, pp. 65–71, 2016.

[70] T. G. Dietterich, "Approximate statistical tests for comparing supervised classification learning algorithms," *Neural computation*, vol. 10, no. 7, pp. 1895–1923, 1998.

[71] S. L. Salzberg, "On comparing classifiers: Pitfalls to avoid and a recommended approach," *Data mining and knowledge discovery*, vol. 1, no. 3, pp. 317–328, 1997.

Gerardo Ibarra-Vazquez Postdoctoral Researcher at the Advanced Artificial Intelligence research group at Tecnológico de Monterrey. Ph.D in Computer Science from the Universidad Autónoma de San Luis Potosí. His research interests are robustness evaluation of image classification models, adversarial attacks in deep learning, deep learning in image processing, evolutionary algorithms for image classification, saliency object detection, and interest point detectors.

María Soledad Ramírez-Montoya PhD in Philosophy and Educational Sciences from the University of Salamanca. She coordinates the interdisciplinary research groups "Scaling Reasoning for Complexity for All" and the Educational Technology research unit, and "Educational Technology Unit", at the Institute for the Future of Education of the Tecnológico de Monterrey (Mexico). She is Chair UNESCO "Open Educational Movement for Latin America" and of the International Council for Open and Distance Education "Open Educational Resources for Latin America". Researcher level 3 of the National System of Researchers of Mexico. Her lines of research are educational innovation and open education. She focuses her activities on the dynamization of educational initiatives, with innovation, research and global sense, as a means of transformation and social impact for lifelong learning and sustainable development.

Mariana Buenestado-Fernández Degree in Psychopedagogy and PhD in Educational Sciences from the University of Cordoba (Spain). She is currently working as Assistant Professor at the Department of Education of the University of Cantabria (Spain), where she carries out teaching and research tasks. Her line of research focuses on teacher training and educational innovation for social responsibility and technological.