Informatics in Education, 2023, Vol. 00, No. 00, –

DOI: 10.15388/infedu.2023.23
ACCEPTED VERSION (WITHOUT THE JOURNAL LAYOUT).

Online Education vs Traditional Education: Analysis of Student Performance in Computer Science using Shapley Additive Explanations

Małgorzata Charytanowicz
Department of Computer Science, Lublin University of Technology, Lublin, Poland
Systems Research Institute, Polish Academy of Sciences, Warsaw, Poland
e-mail addresses: m.charytanowicz@pollub.pl

Received: October 2022

Abstract. Nowadays, the rapid development of ICT has brought more flexible forms that push the boundaries of classic teaching methodology. This paper is an analysis of online teaching and learning forced by the COVID-19 pandemic, as compared with traditional education approaches. In this regard, we assessed the performance of students studying in the face-to-face, online and hybrid mode for an engineering degree in Computer Science at the Lublin University of Technology during the years 2019-2022. A total of 1827 final test scores were examined using machine learning models and the Shapley additive explanations method. The results show an average increase in performance on final tests scores for students using online and hybrid modes, but the difference did not exceed 10% of the point maximum. Moreover, the students' work had a much higher impact on the final test scores than did the study system and their profile features.

Keywords: higher education, online learning, XGBoost, SHAP values, COVID-19, student motivation.

1. Introduction

Higher education has undergone tremendous changes during the COVID-19 pandemic. The onset of the pandemic, falling in the summer semester of 2020, forced students to stay at home and use online learning platforms. The sudden transition to online education exposed, however, some challenges, as well as benefits. For the 2020/2021 and 2021/2022 academic years, different action strategies were then adopted: some universities reinstated in-person participation in classes, some remained fully online and others applied hybrid solutions.

Online learning is made possible due to harnessing the potential of the Internet and independent technological devices to develop and to allow the use of on-line materials and tutoring for education purposes, instructional delivery and management of programs through networked interactivity (Fry, 2001; Zhang and Wu, 2022). There are two types of online learning now proffered: asynchronous online learning and synchronous online learning (Hrastinski, 2008).

Asynchronous learning provides more flexible time for students and teachers since they need not be online at the same time. Students can review the learning materials at any time and spend more time in forums, virtual libraries, sites accessing lecturers’ online notes or online discussions boards. Unfortunately, asynchronous learning has several disadvantages. Among such are the lack of in-person contact with the lecturer and the feeling of isolation and loneliness due to the absence of interpersonal contacts.

Synchronous online learning refers to real-time online learning that enables students and teachers to interact at the same time (Hong et al., 2020; Paul and Jefferson, 2019; Rawat and Singh, 2020). It provides a live platform that allows more direct interaction and immediate response between teachers and students. The main strength of synchronous online learning lies its immediately and efficiently communicated environment (Giesbers et al., 2014).

Along with several benefits and its flexibility, online learning also has several disadvantages. Among such are the lack of face-to-face interactions resulting in the feelings of isolation and loneliness (Stewart and Lowenthal, 2022). Many researchers (Hong et al., 2020; Paul and Jefferson, 2019; Rawat and Singh, 2020) have clearly shown that in-person participation has the greatest impact on student motivation - which is a prerequisite for successful learning. For online-only students, satisfaction with their studies is halved, with the greatest negative impact being in the areas of academic and personal development. Motivation to learn is a phenomenon largely conditioned by individual traits, particularly, the emotional factors of the individual. In online learning, it is easy to give in to external influences and distractions (Ha and Wong, 2010; Ditta, 2020; Salguero et al., 2021).

Several research studies have been carried out recently to explore the effectiveness of online education in comparison to traditional offline education (Cotero et al., 2020; Harwood et al., 2018; Kumar et al., 2009). On-line educational platforms, for instance, seem to be very helpful in furthering individual modes of work by making self-studies more efficient (Charytanowicz, 2019; Núñez et al., 2017; Suszyński et al., 2020). Due to changes caused by the COVID-19 pandemic, online education has been popularized and more fully developed. The pandemic has increased not only the importance of emergency online studying, but has also provided an opportunity for a broader discussion of the efficacy of online higher education (Makarova, 2021; Singh et al., 2021; Stewart and Lowenthal, 2022).

Large data resources generated by online educational applications have allowed the usage of machine learning models to explore students' performances. The prediction output generated by machine learning models can be explained by applying Shapley analysis. The work (Sahlaoui et al., 2021) can serve as a good example of this. This paper presents preliminary research depicting the importance of tree-based machine learning algorithms and their application in predicting student performance. The results showed that the use of the proposed strategies and techniques improved the accuracy of the prediction models. Moreover, the SHAP value and the associated visualizations increased transparency.

Through machine learning methods and the SHAP approach, certain researchers have aimed to discover which features have the most significant impact on ADHD student performance in arithmetic, writing and reading (Balbino et al., 2022). Here, SHAP values aided in identifying and assessing the most relevant features for academic performance.

However, currently there is a lack of papers that, with the aid of Shapley analysis, allow for better explanation and understanding of student performance. Therefore, the study conducted in this paper investigates the problem of student performance forecasting from the point of view of input features. Our research contributes to a proper assessment of online learning effects by comparing the relationships between student learning in the traditional classroom (before the pandemic) with synchronous fully online learning due to the pandemic and the hybrid mode introduced in order to adapt to the ongoing pandemic. For addressing the assessment of the student performance, our investigation deals with two basic courses in the CS area: Introduction to Computer Science; Numerical Analysis Algorithms. We have attempted to answer the following two main research questions:

1. What is the effect of online teaching and learning forced by COVID-19 pandemic on student performance of the 1st degree studies in Computer Science?
2. What are the main factors that affect student knowledge?

The remainder of the paper is organized as follows. Section 2 presents the context of the study, materials and methodology followed so as to discuss the fundamentals of feature importance realized by Shapley analysis. Section 3 explores and explains the results obtained. Section 4 concludes our work, with limitations and future scope presented in Section 5.

2. Materials and methods

2.1. *Context of the study*

The study was carried out among students of undergraduate engineer studies in CS at the Lublin University of Technology (LUT) in Poland, during the winter semesters (October – mid-February) of the academic years 2019/2020, 2020/2021 and 2021/2022. It should be noted that LUT has implemented a quality assurance system of education in accordance with the requirements of the Ministry of Higher Education and Science, and the Faculty of Electrical Engineering and Computer Science has received positive accreditation and has been assigned a very high scientific category (A). The curricula created for Computer Science covers all areas of knowledge described according to the ACM/IEEE undergraduate education model (Miłosz et al., 2014; Tang et al., 2018). This curricula was modified in the academic year 2019/2020 based on ACM/IEEE and ACM CCECC documentation. The ACM CCECC document ("Computer Science Curricular Guidance", 2018) defines the following knowledge areas:

(1) Algorithms and Complexity
(2) Computational Science
(3) Discrete Structures
(4) Human-Computer Interaction
(5) Networking and Communications
(6) Parallel and Distributed Computing
(7) Programming Languages
(8) Software Engineering
(9) Social Issues and Professional Practice
(10) Architecture and Organization
(11) Cybersecurity
(12) Graphics and Visualization
(13) Information Management
(14) Operating Systems

(15) Platform-based Development
(16) Software Development Fundamentals
(17) Systems Fundamentals.

All these areas are included in the currently implemented curricula and they are completed with measurable student learning outcomes. Moreover, the LUT cooperates with several ICT companies in order to adapt its curricula to the needs of local and international labour markets and changes in ICT technology. Both full-time and part-time study options are offered and they last 3.5 years.

During the winter semester of the academic year 2019/2020, the teaching was based on traditional face-to-face lectures and laboratories with smaller groups of about 15 students. The educational process was aided by a rich Moodle learning platform.

From the summer semester of the academic year 2019/2020 until the summer semester of the academic year 2020/2021, all classes were conducted using synchronous online teaching and learning due to the SARS-CoV-2 pandemic and the suspension of classroom-based studies. The decision to transition to online learning was made 15th March of 2020. The use of online education platforms hence had become a necessity. Face-to-face classes were replaced by fully online teaching sessions. Synchronous learning was similar to traditional classroom approaches. Classes were taught in their normal schedule and students were required to log in and participate in classes by connecting by way of their personal computing devices, with cameras operative. All activities including presentations, student work, discussions and feedback were done completely online, while physical presence was replaced by screen sharing and virtual communication via MS Teams. The educational strategy aided by the Moodle learning platform was maintained.

In the academic year 2021/2022, the hybrid mode was introduced due to the need to adapt to the ongoing epidemic. Laboratories were conducted as face-to-face classes, while lectures were replaced by synchronous online teaching sessions and screen sharing via the MS Teams platform because of the very large number of students and increased risk of infection for COVID-19. As before, teaching and learning was aided by the Moodle platform.

2.2. *Course selection*

To compare the effectiveness of traditional, online and hybrid education on student performance, the courses were selected according to three main criteria: (1) the courses covered algorithms and programming; (2) the course curricula was not modified in the academic years 2019/2020, 2020/2021 and 2021/2022; (3) student knowledge was assessed according to the same tools using Moodle.

Two compulsory courses of the 1st and 3rd semester of the undergraduate engineer studies were selected: Introduction to Computer Science, and Numerical Analysis Algorithms. These are basic courses in the CS area, and have a four-credit workload and a five-credit workload, respectively, according to the European credit transfer system. In the full-time study system curriculum, both courses last 15 weeks, including two hours of lecture and two hours of laboratory per week – in total, 60 hours per course. Lectures and classes are held from Monday to Friday. The equivalent part-time study system covers exactly the same curricula and learning outcomes, albeit with the number of face-to-face teaching hours reduced to half. Lectures and classes are held two or three

times a month on weekends: Saturday-Sunday 08:00-20:30. This study system allows the student to work more hours, or to pick up a full-time job. Because of this loss of class-room hours, students are expected to spend much more hours in self-study.
The first course covers core computer science and programming concepts, as well as the development, description, efficiency analysis and correctness of basic algorithms, with basic practical coding skills developed through Python (Charytanowicz et. al., 2020). The second course is offered by most curricula in CS as a core course. It includes the learning of fundamental techniques for deriving efficient numerical solutions to problems in science and engineering. These are: errors in numerical computations, interpolation, approximation of functions, integration, differential equations, direct and iterative methods in linear algebra. The intent is for students to be able to self-design algorithms and code programs in C++ language.

2.3. *Study participants*

The study participants were selected from students of undergraduate engineer studies in CS of the 1st and 3rd semester (sem.) in the academic years 2019/2020, 2020/2021 and 2021/2022. Students were invited to complete the short online quizzes and final test via Moodle platform of two basic courses in the CS area: Introduction to Computer Science (1st sem.) and Numerical Analysis Algorithms (3rd sem.). Positive response was given by more than 80% of all students of each year and semester, and ranged between 81% and 93%. Table 1 shows the number of enrolled students with regard to particular year and semester. The last two rows contain the number and percentage of these students participating in the quizzes and tests. They naturally formed the study group of participants aged 18 to 22 years.

Table 1
Number of students enrolled at the university and number of students who participated in the quizzes and tests of the particular academic year and semester

Number of students	2019/2020 Face-to-face mode		2020/2021 Fully online mode		2021/2022 Hybrid mode	
	1st sem.	3rd sem.	1st sem.	3rd sem.	1st sem.	3rd sem.
enrolled at the university	414	263	416	306	329	339
participating in the quizzes and tests	334	230	374	285	288	316
Percentage of participants	81%	87%	90%	93%	88%	93%

A total of 1827 final test scores were collected. Table 2 lists the number of participants assigned to particular academic year, semester and study system, as well as their gender.

Table 2
Number of participants assigned to particular academic year, semester and study system – and their gender

Study system	2019/2020 Face-to-face mode		2020/2021 Fully online mode		2021/2022 Hybrid mode	
	1st sem.	3rd sem.	1st sem.	3rd sem.	1st sem.	3rd sem.

Full-time (male / female)	283 (251 / 32)	194 (172 / 22)	316 (275 / 41)	234 (207 / 27)	226 (192 / 34)	279 (237 / 42)
Part-time (male / female)	51 (47 / 4)	36 (34 / 2)	58 (54 / 4)	51 (47 / 4)	62 (57 / 5)	37 (35 / 2)
Total	334	230	374	285	288	316

In total, males constituted 88% of the study sample. Regarding nationality, 86% came from Poland and 14% from other countries, mainly Ukraine.

2.4. *Procedure*

For both selected courses, appropriate Moodle courses were created to aid the teaching and learning process. Each student was given a personal log-in name and password to access the Moodle course. The Moodle course consisted of proffered resources and activities that were to be completed. This was used not only as a repository of materials, but also as an interactive learning tool for motivating students to work and for assessing student progress. An additional strategy was introduced to help students self-check for common errors. After each issue topic or lecture, students completed a short online quiz, which was set to have a time limit and to allow only one attempt with a random sequential navigation mode, one question per page. Immediate feedback was given to students, indicating which questions were answered correctly. The total score for these quizzes was 20. Tests were developed by the lecturer that created the curricula and course material and had prepared and delivered the lecture.

It is worth noting that during the academic years 2019/2020, 2020/2021 and 2021/2022, both course lectures were conducted by the same individuals. Lecture presentations and additional materials were prepared in 2019 year according to the modified curricula and they did not change during the conducted research. Test questions were modified each year to prevent their sharing between students.

Student knowledge was assessed throughout the semester via a final online test worth over all, 20 points (1 point for each question), performed through Moodle. The final test included randomly selected multiple-choice, numerical, short answer or essay questions related to the issues discussed during the course. The sequential navigation mode was also applied. In this, students cannot return to previously attempted questions. Large groups of students were divided into smaller groups and all students took the final test simultaneously to minimize the possibility of answers being exchanged among students.

Despite the sudden transition to online education in March 2020, synchronous online teaching and learning allowed conducting teaching and learning activity according to the implemented curricula and learning outcomes. Moreover, the short online quizzes and final tests performed through Moodle were appropriate to the online mode. The main difference was in the personal contact with the teacher. In the winter semester of the academic year 2019/2020, classes were conducted face-to-face and final tests were held at the University. When the online mode was enacted, final tests were conducted through online proctoring.

2.5. *Extreme gradient boosting*

Extreme gradient boosting (XGBoost) was built as an enhanced version of the gradient boosting decision tree (GBDT) algorithm (Chen and Guestrin, 2016). XGBoost provides strong regularisation by adopting a stepwise shrinkage process instead of the traditional weighting process provided by GBDT. It also optimizes the loss function using first and second-order gradient statistics.

For a given data set $D = \{(\mathrm{x}_i, y_i)\}$ with n examples and m features, a tree ensemble model uses K additive functions $f_k \in \mathcal{F}$ to predict the output values:

$$\hat{y}_i \;= \sum_{k=1}^{K} f_k(\mathrm{x}_i) \ ,$$

where: $\mathcal{F} = \{f(x) = w_{q(\mathrm{x})}\}$ is the space of regression trees, $w \in \mathbb{R}^T$ denotes leaf weights, and $q: \mathbb{R}^m \to T$ represents the structure of each tree that maps an example to the corresponding leaf index. Each f_k corresponds to an independent tree structure q and leaf weights w. To reduce errors within the ensemble trees, the objective function of XGBoost can be represented as:

$$\mathcal{L}^{(t)} = \sum_{i=1}^{n} l\left(y_i, \hat{y}_i^{(t-1)} + f_t(\mathrm{x}_i)\right) + \Omega(f_t) \ ,$$

where: t shows the repetitions in order to minimize the errors and the term Ω penalizes the complexity of the regression tree functions. This can be expressed as:

$$\Omega(f_k) = \gamma T + \frac{1}{2}\lambda \sum_{i=1}^{T} w_i^2 \ .$$

Here, a differentiable convex loss function l measures the difference between the prediction $\hat{y}$ and the target y_i, T is the number of leaf nodes, while the parameters γ and λ are used to control the number of leaf nodes and the weight of leaf nodes. Second-order approximation is applied to optimize the objective function. This can be simplified as follows:

$$\tilde{\mathcal{L}}^{(t)} = \sum_{i=1}^{n}\left(g_i f_t(\mathrm{x}_i) + \frac{1}{2} h_i f_t^2(\mathrm{x}_i)\right) + \gamma T + \frac{1}{2}\lambda \sum_{j=1}^{T} w_j^2 \ ,$$

where: g_i and h_i are the first and second order statistics on the loss function l and $I_j = \{i | q(\mathrm{x}_i) = j\}$ denotes the instance set of leaf j. For a fixed structure $q(\mathrm{x})$, the solution expression is given as

$$w_i^* = -\frac{\sum_{i \in I_j} g_i}{\sum_{i \in I_j} h_i + \lambda} \ ,$$

$$\tilde{\mathcal{L}}^{(t)}(q) = \sum_{j=1}^{T} \frac{\left(\sum_{i \in I_j} g_i\right)^2}{\sum_{i \in I_j} h_i + \lambda} + \gamma T \ ,$$

where: w_i^* represents the optimal weight of leaf j, while $\tilde{\mathcal{L}}^{(t)}$ denotes the corresponding optimal value and can be used to measure the quality of a tree structure q.

2.6. *Shapley additive explanations*

Machine learning models are useful predictive tools that are, due to their high efficiency and flexibility, frequently applied in regression tasks. To evaluate the performance of the machine learning models, mean squared error (MSE), root mean squared error (RMSE), mean absolute error (MAE) and coefficient of determination (R-squared) are employed for estimating the accuracy of the forecast results. The formulae are as follows:

$$\mathrm{R}^2 = 1 - \frac{\sum_{i=0}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=0}^{n}(y_i - \bar{y})^2} \ ,$$

$$\mathrm{MSE} = \frac{1}{n}\sum_{i=0}^{n}(y_i - \hat{y}_i)^2 \ ,$$
$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=0}^{n}(y_i - \hat{y}_i)^2} \ ,$$
$$\mathrm{MAE} = \frac{1}{n}\sum_{i=0}^{n}|y_i - \hat{y}_i| \ ,$$

where: n is the sample size, $\hat{y}_i$ is the i-th prediction result of the model, y_i is the i-th observed value and $\bar{y}$ denotes the average of y_i, $i = 1, 2, \dots, n$. The lower value of MSE, RMSE and MAE implies higher accuracy of a regression model. The value of R-squared ranges from 0 to 1 and is interpreted as a percentage. A higher value of R-squared is considered desirable.

Explaining the model and understanding how the features are related to the outputs can be done using the Shapley additive explanations method (SHAP) (Aas et al., 2021; Friedman, 1986; Lundberg and Lee, 2017). The Shapley values method is a mathematical concept based on cooperative game theory. The idea used for explanations of model predictions is to treat features that explain the prediction models as players and the prediction as the total payout. This method requires retraining the model on all coalitions (subsets) of players $S \subseteq F$, where F is the set of all players. It assigns an importance value to each player i that represents the effect on the model prediction of including that player. To compute this effect, a model $f_{S\cup\{i\}}$ is trained with that player present, and another model f_S is trained with the player withheld. The importance or the influence of the player is obtained by calculating the difference $f_{S\cup\{i\}}(x_{S\cup\{i\}}) - f_S(x_S)$ of model prediction with and without the player i. Since the effect of withholding a player depends on other players in the model, the preceding differences are computed for all possible coalitions $S \subseteq F\backslash\{i\}$. The Shapley values are then computed as weighted average of all contributions of a player i to a model score, as follows:

$$\phi_i = \sum_{S\subseteq F\backslash\{i\}} \frac{|S|!(|F|-|S|-1)!}{|F|!}\left(f_{S\cup\{i\}}(x_{S\cup\{i\}}) - f_S(x_S)\right) \ ,$$

where: $|A|$ refers to the cardinal of the set A and x_S represents the values of the input players in the coalition S.

3. Results

The research procedure consisting of two steps. In order to analyze student performance of the 1st degree studies in Computer Science, in both traditional, online and hybrid education, we applied the mean, standard deviation (SD) and 95% confidence interval (CI) for the tests scores of two courses: Introduction to Computer Science (1st semester) and Numerical Analysis Algorithms (3rd semester). The following study features were considered: study mode (face-to-face, fully online and hybrid) and study system (part-time studies and full-time studies).

Features affected student knowledge were evaluated using an efficient and flexible machine learning algorithm called 'extreme gradient boosting regression' (XGBoost). XGBoost was built as an enhanced version of the gradient boosting decision tree algorithm (Wade, 2020; Freeman et al., 2016). The interpretability of the model and feature analysis was done by the Shapley additive explanations method (Lundberg and Lee, 2017).

3.1. *Comparative analysis of the final test scores*

Table 3 and Table 4 present the mean, SD and 95% CI of the final test scores for face-to-face, online and hybrid modes of full-time and part-time studies. There were no statistically significant differences between scores of male and female students and therefore the results were presented for the whole group. In the tables, the bold font indicates best and worst results.

Table 3
Mean, SD and 95% CI of the final test scores of the 1st semester course: Introduction to Computer Science

Study system	Study mode					
	Face-to-face 2019/2020		Fully online 2020/2021		Hybrid 2021/2022	
	Mean ± SD	95% CI	Mean ± SD	95% CI	Mean ± SD	95% CI
Full-time	**11.01 ± 3.94**	10.54 – 11.46	**13.70 ± 2.99**	13.37 – 14.04	13.48 ± 2.90	13.10 – 13.86
Part-time	12.13 ± 4.32	10.91 – 13.34	11.74 ± 3.11	10.92 – 12.56	13.67 ± 3.09	12.88 – 14.45

As can be seen in Table 3, higher final scores of the course: Introduction to Computer Science were achieved for the fully on-line manner of full-time studies (13.70 ± 2.29). These scores are also similar to the results obtained for the hybrid mode. However, significantly worse results can be seen for the face-to-face manner of full-time studies (11.01 ± 3.94), the difference is about 2 points.

Table 4
Mean, SD and 95% CI of the final test scores of the 3rd semester course: Numerical Analysis Algorithms

Study system	Study mode					
	Face-to-face 2019/2020		Fully online 2020/2021		Hybrid 2021/2022	
	Mean ± SD	95% CI	Mean ± SD	95% CI	Mean ± SD	95% CI
Full-time	**10.64 ± 3.25**	10.18 – 11.10	**12.37 ± 3.27**	11.95 – 12.79	12.13 ± 3.10	11.76 – 12.49
Part-time	11.42 ± 3.10	10.37 – 12.47	11.31 ± 2.65	10.57 – 12.06	12.03 ± 3.59	10.83 – 13.23

Table 4 allows us to draw similar conclusions – higher final scores of the course: Numerical Analysis Algorithms were achieved for the fully online manner of full-time studies (12.37 ± 3.27). In contrast, significantly worse results can be seen for the face-to-face manner of full-time studies (10.64 ± 3.25) and the difference is about 2 points.
Table 5 and Table 6 present the Mean, SD and 95% CI of the short quizzes scores for face-to-face, online and hybrid modes of full-time and part-time studies of the 1st semester and the 3rd semester course, respectively. In the tables, the bold font indicates best and worst results.

Table 5
Mean, SD and 95% CI of the short quizzes scores of the 1st semester course:
Introduction to Computer Science

Study system	Study mode: Face-to-face (n = 334)		Fully online (n = 374)		Hybrid (n = 288)	
	Mean ± SD	95% CI	Mean ± SD	95% CI	Mean ± SD	95% CI
Full-time	13.42 ± 3.49	13.01 – 13.83	14.89 ± 2.81	14.58 – 15.20	15.77 ± 2.32	15.47 – 16.07
Part-time	**16.08 ± 3.25**	15.16 – 16.99	**13.07 ± 2.83**	12.32 – 13.81	15.23 ± 3.33	14.38 – 16.07

Table 6
Mean, SD and 95% CI of the short quizzes scores of the 3rd semester course:
Numerical Analysis Algorithms

Study system	Study mode: Face-to-face (n = 230)		Fully online (n = 285)		Hybrid (n = 316)	
	Mean ± SD	95% CI	Mean ± SD	95% CI	Mean ± SD	95% CI
Full-time	**14.34 ± 3.31**	13.87 – 14.81	14.40 ± 3.05	14.01 – 14.80	14.94 ± 3.00	14.59 – 15.29
Part-time	14.67 ± 3.02	13.64 – 15.69	**15.41 ± 2.24**	14.78 – 16.04	15.14 ± 2.56	14.28 – 15.99

The student scores of final tests were strongly positively correlated with the short quizzes scores for both courses: Introduction to Computer Science (the Pearson correlation coefficient $r = 0.8038$) and Numerical Analysis Algorithms ($r = 0.7978$).

3.2. *Machine learning models*

The following independent variables were considered:

- short quizzes score (0 – 20 points);
- study mode (face-to-face, fully online, hybrid);
- course selection (Introduction to Computer Science – 1st semester, Numerical Analysis Algorithms – 3rd semester);
- study system (part-time, full-time);
- nationality (Polish, others);
- gender (woman, man).

The dependent variable (outcome) is the final test score (0 – 20 points).

3.2.1. *Model development*

With regard to examining the prediction performance of the XGBoost regression model, the input data was divided into training and testing subsets composed of 80% and 20% of samples, respectively. To improve the precision, five-folds cross validation was applied to train and evaluate our models. The XGBoost parameters were optimized using a simple grid search algorithm (Pedregosa et al., 2011) to estimate the optimal parameters. Within the process of validation, the influence of the following parameters was explored:

- learning rate: 0.01, 0.1;
- the maximum depth of a tree: 3, 5, 7, 9, 15;
- the minimum sum of weights of all observations required in a child: 1, 2, 4, 6;
- the fraction of observations to be used as random samples for each tree: 0.5, 0.7;
- subsample ratio of columns when constructing each tree: 0.5, 0.7;
- the number of trees to fit: 100, 250, 500, 1000.

3.2.2. *Model validation*

After performing a five-fold cross validation, the optimal, in terms of the highest prediction ability, set of XGBoost parameters was obtained and used to verify the model on the independent test set. The R-squared score was utilized to evaluate the performance of the model in comparison to the classical multiple regression model. We found that the XGBoost regression model outperformed the multiple regression model in both training and testing, with the average R-squared values being above 70%. We also noted that the XGBoost model obtained higher R-squared values with regard to all five folds. In addition, the XGBoost model resulted in more consistent values and smaller MSE, RMSE and MAE values, compared to the multiple regression model (see Table 7 and Table 8).

Table 7
The average evaluation metrics in the training set for the five folds

Model	MSE	RMSE	MAE	R-squared
XGBoost regression	3.034	1.742	1.254	0.744
Multiple regression	4.034	2.008	1.470	0.660

Table 8
The average evaluation metrics in the test set for the five folds

Model	MSE	RMSE	MAE	R-squared
XGBoost regression	3.550	1.884	1.302	0.701
Multiple regression	4.088	2.011	1.479	0.655

The difference between the original and predicted final test scores is presented in Figure 1. The proposed XGBoost model offered a balanced prediction throughout the training and testing data sets.

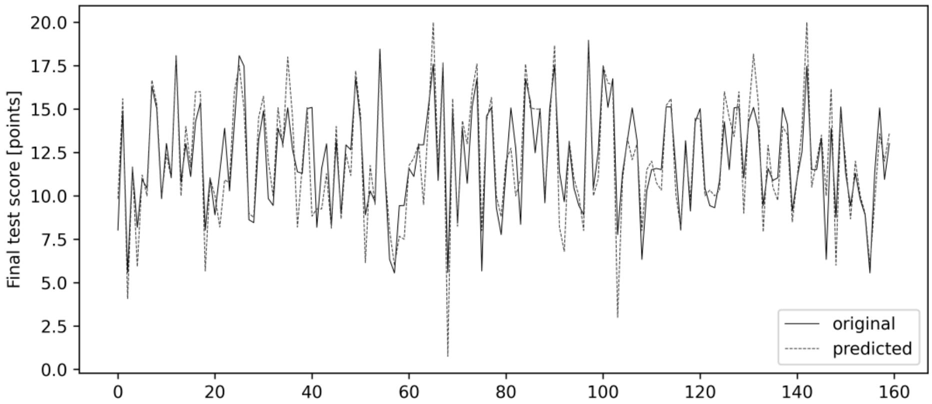

Fig. 1. The final test score prediction results according to the XGBoost model for exemplary 160 testing set participants.

While the model attained good accuracy rates, it is difficult to comprehend. Still, thanks to the SHAP values, this can be explained, hence it can provide both global and local interpretation.

3.2.3. *Model interpretation*

The feature importance of the models is plotted in a bar chart using the SHAP aggregations, as illustrated in Figure 2.

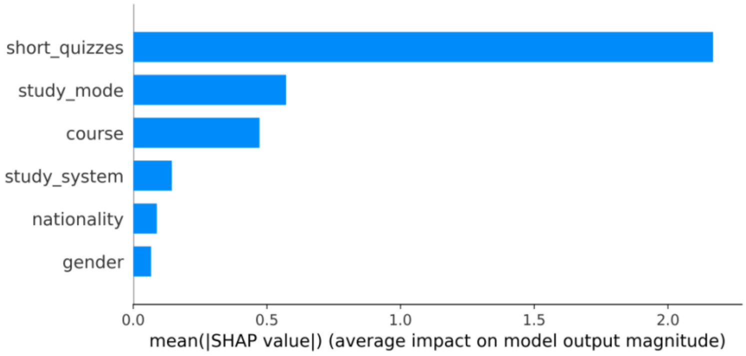

Fig. 2. Global feature importance for the XGBoost model based on the mean absolute magnitude of the SHAP values.

A SHAP value being much closer to zero means that the data point contributes very little to the predictions. In contrast, if the SHAP value is strongly positive or strongly negative, this outcome reveals that the data point greatly contributes towards predicting the positive or negative class.

Figure 2 highlights the major importance of short quizzes scores, in that the large positive/negative SHAP values indicate that the change of this feature can have a more noticeable influence than that of the other variables. Short quizzes score (2.169), study mode (0.572), course (0.473) and study system (0.145) display the highest SHAP values. As positive short quizzes score SHAP values were evident, we can conclude that if student short quizzes scores are higher, then student performance sees improvement accordingly. However, study profile features such as study mode, course (semester) and study system were also important with regard to the student performance. By contrast, student profile features such as nationality (0.089) and gender (0.067) did not have a huge effect on student performance, indicating that the change of these features does not have a noticeable influence on model prediction.

Through SHAP, we created summary plots (Figure 3) to demonstrate the feature importance and contribution in predictions. Herein, all of the features are listed on the y-axis in rank order, the top one being the greatest contributor to the predictions and the bottom one being the least or zero-contributor. The x-axis depicts the Shapley value, and the color shows the degree of the effect (the red and blue colors show positive and negative predictions, respectively).

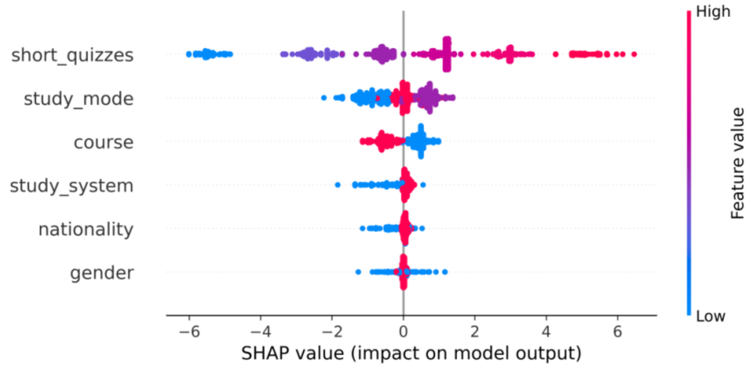

Fig. 3. Summary plots for the XGBoost model.

Figure 3 reveals that if student short quizzes scores are higher, then student performance shows improvement accordingly. This means that student's work and commitment have a positive impact on the final test scores. The second row of the summary plot relates to the traditional face-to-face, fully online and hybrid study mode. In terms of the study mode, a negative effect on student performance is found for the face-to-face mode and a positive effect is found for the fully online mode. This effect is much lower in comparison to short quizzes scores. This finding is consistent with our statistical analysis. The final grades of the first course: Introduction to Computer Science are higher than for the second course: Numerical Analysis Algorithms. Furthermore, the full-time study system is positively related to student performance, while Nationality and gender do not have a huge effect on student performance. However, we can notice that Polish students achieved a bit better results.

This plot is a suitable tool for obtaining an improved understanding of how certain features affect the model decision. Still, to obtain a deeper understanding of our model, we then developed related force plots. These provided us with information on feature contributions for specific observations. In our study, the force plot in Figure 4 highlights the features responsible for predicting student performance. Features that have more predictive control are shown in red, whereas features that have lower predictive control are shown in blue. Moreover, the output value is the prediction with features, whereas the base value is the value that would be predicted without any features, that is, the mean prediction.

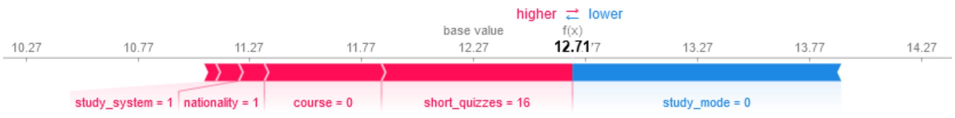

Fig. 4. Force plot depicting feature contribution towards a single prediction.

Figure 4 shows that the baseline is 12.27 and the actual prediction is 12.71. We found that study system, nationality, short quizzes score and course features can increase the prediction value, while study mode can decrease the final output prediction. The utilized force plots, therefore provided information about the key features responsible for student performance at the observation level.

4. Final comments and summary

The COVID-19 lockdown coincided with the implementation of modified curricula in the teaching of Computer Science (CS) in LUT, in 2019. This allowed obtaining data to ascertain differences in learning success for face-to-face, synchronous online and hybrid learning modes during the years 2019-2022. All students followed the same study curricula consistent with the expectations of the labour market (Łukasik et al., 2020; Miłosz et al., 2020). It is worth adding that this curricula covers items that are of the greatest demand in the ICT industry.

We considered XGBoost for regression and SHAP values to be powerful tools for student performance prediction and interpretation. The conducted studies formulate important conclusions that explain student-learning achievement. According to SHAP values, short quizzes scores had the greatest impact on the student performance. This means that student's effort and commitment are key elements to being successful in the final test scores. The second most important factor is the study mode: traditional fully on line, synchronous online and hybrid. Despite concerns over the COVID-19 lockdown and the sudden transition to online learning, final test scores of students using synchronous online and hybrid learning increased by about 10% of the point maximum. However, the study profile features such as course (semester) were also important with regard to student performance. According to the SHAP values, the study system had a bit lower impact on the final test score. In contrast, student profile features of nationality and gender had the lowest impact on final test scores

The conducted analysis revealed a significant increase in performance on final tests scores for students using the synchronous online mode and the hybrid mode, in comparison to traditional face-to-face studies. The difference was about 2 points regardless of the study semester (see Table 3 and Table 4, Section 3.1). Furthermore, in the case of traditional studies, in the 1st and 3rd semester, higher grades were obtained in the part-time study system when compared to the full-time system. In contrast, in the case of synchronous online studies, in the 1st and 3rd semester, higher grades were obtained in the full-time study system. However, these differences are not statistically significant and they were of about 1 point. In the case of hybrid studies, we noticed similar levels of student performance.

Summarizing, the results obtained substantiate the use of online synchronous learning as a form of learning that supports traditional educational approaches and enables university education to be more affordable and accessible.

5. Limitations and future work

Certain limitations of this work result from the act that the research concerned Computer Science students of undergraduate engineer studies who had to switch to the online learning due to the COVID-19 lockdown. Therefore, we examined the effectiveness of traditional and fully online and hybrid education on student performance in Computer Science, but we did not investigate the effects of the three modes of learning within the whole University. Still, it is important to emphasize that the research concerned all students of the first two years of studies, both full-time and part-time systems. To full understand the effect of mode of learning, different machine learning prediction models and other explainable approaches to investigating the problem need to be considered and applied. This is the task to be included in future works. We will also focus on the extending the dataset by using more variables and by combining the effects of several variables. Additionally, further work will concern the complete cycle of studies. The main aim will be to investigate the influence of online learning on the final results of Computer Science students within the undergraduate engineer program. We also will focus on the hybrid education system that was initiated in the 2021/2022 academic year and will most likely be introduced permanently in the part-time studies curricula.

References

Aas, K., Jullum, M., Loland, A. (2021). Explaining individual predictions when features are dependent: More accurate approximations to Shapley values. *Artificial Intelligence*, 298, 1–24. https://doi.org/10.1016/j.artint.2021.103502.

ACM Committee for Computing Education in Community Colleges. (2018). *Computer Science Curricular Guidance for Associate-Degree Transfer Programs with Infused Cybersecurity*. ACM, New York.

Balbino, M., Jandre, C., De Miranda, D., Nobre, C. (2022). Predictions of academic performance of children and adolescents with ADHD using the SHAP approach doi:10.3233/SHTI22015.Charytanowicz, M. (2019). E-learning technologies in lifelong learning: opportunities and challenges. In: *ICERI'2019 Proceedings: 12th International Conference of Education, Research and Innovation*, 9414–9421.

Charytanowicz, M., Skublewska-Paszkowska, M., Miłosz, M., Łukasik, E. (2020). What programming language should be taught to computer science students? – that is the question. In: *INTED 2020: 14th International Technology, Education and Development Conference*, 747–756.

Chen, T., Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. In Proceedings of the 22nd 611 ACM SIGKDD International Conference on Knowledge

Discovery and Data Mining - KDD’16, 612, 785–794.
Cotero, K., Rodríguez, L., Aguilar, A. (2020). Appropriation of skills in students who migrated from traditional education model to an online education model, derivated from COVID-19. In: *2020 X International Conference on Virtual Campus*, 1–4. https://10.1109/JICV51605. 2020.9375801.
Ditta, A.S., Strickland-Hughes, C.M., Cheung, C.S., Wu R. (2020). Exposure to information increases motivation to learn more. *Learning and Motivation*, 72, 101668. https://10.1016/j.lmot.2020.101668.
Freeman, E. A., Moisen, G. G., Coulston, J. W., Wilson, B. T. (2016). Random forests and stochastic gradient boosting for predicting tree canopy cover: comparing tuning processes and model performance, *Canadian Journal of Forest Research*, 46(3), 323–339.
Friedman, J.W. (1986). *Game Theory with Applications to Economics*. Oxford University Press.
Fry, K. (2001). E-learning markets and providers: Some issues and prospects. *Education + Training*, 43(4), 233–239. https://doi.org/10.1108/EUM0000000005484.
Giesbers, B., Rienties, B., Tempelaar, D., Gijselaers, W. (2014). A dynamic analysis of the interplay between asynchronous and synchronous communication in online learning: The impact of motivation. *Journal of Computer Assisted Learning*, 30(1), 30–50. https://doi.org/10.1111/jcal.12020.
Ha S., Wong S. (2010). Motivating students to learn through good and helpful coursework feedback. *Education 2010: Inspiring the Next Generation of Engineers*, 1-13. https://www.researchgate.net/publication/46284803.
Harwood, K.J., McDonald, P.L. Butler J.T., Drago, D., Schlumpf, K. S. (2018). Comparing student outcomes in traditional vs intensive, online graduate programs in health professional education. *BMC Medical Education*, 18(240), 1-9. https://10.1186/s12909-018-1343-7.
Hong, Y., Li, X., Lin, Y., Xie, J., Yan, X., Lin Z. (2020). A Comparative Study of Online Education and Traditional Offline Education During COVID-19. *Research Square*, 1-19. https://10.21203/ rs.3.rs-61593/v1.
Hrastinski, S. (2008). Asynchronous & synchronous e-learning. *Educause Quarterly*, 31(4), 51–55.
Kumar, S.A., Amaluddin, Y.M., Win, O.W. (2009). A Comparative Study between Traditional Learning and E-Learning. In T L Centre, CSM (Ed.), *Teaching and Learning Open Forum*, 1-7.
Lundberg, S.M., Lee, S.-I. (2017). A Unified Approach to Interpreting Model Predictions. *Advances in Neural Information Processing Systems*, 30, 4765–4774.
Łukasik, E., Miłosz, M., Skublewska-Paszkowska, M., Charytanowicz, M. (2020). Adjusting the undergraduate computer science curricula to labour market requirements based on the ACM/IEEE model. In: *INTED 2020: 14th International Technology, Education and Development Conference*, 739-746.
Makarova, E. (2021). Effectiveness of traditional and online learning: comparative analysis from the student perspective. In: *SHS Web of Conferences 99*, 01019. https:// 10.1051/shsconf/20219901019.
Miłosz, M., Lujan-Mora, S., Bardou, D., Mercelon, A., Penafiel M. (2014). Comparison of Existing Computing Curricula and the New ACM-IEEE Computing Curricula 2013. In: *Proceedings of the 6th International Conference on Education and New Learning*

Technologies, 5808–5818.
Miłosz, M., Charytanowicz, M., Skublewska-Paszkowska, M. (2020). Analysis of trends in ICT and its implication in higher education curricula. In: *INTED 2020: 14th International Technology, Education and Development Conference*, 500–509.
Núñez, J.L.M., Caro, E.T., González, J.R.H. (2017). From Higher Education to Open Education: Challenges in the Transformation of an Online Traditional Course. *IEEE Transactions on Education*, 60(2), 134-142.
Paul, J., Jefferson, F. (2019). A Comparative Analysis of Student Performance in an Online vs. Face-to-Face Environmental Science Course From 2009 to 2016. *Frontiers in Computer Science*, 1(7), 1-9. https://10.3389/fcomp.2019.00007.
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. (2011). *Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research*, 12, 2825-2830.
Rawat, R., Singh, P. (2020). A Comparative Study between Traditional and Online Teaching-Learning: Medical Students' Perspective in the Wake of Corona Pandemic. *National Journal of Community Medicine*, 11(9), 341-345. https://10.5455/njcm. 20200902070715.
Sahlaoui, H., El Arbi, A.A., Nayyar, A., Agoujil, S., Jaber, M.M., et al. (2021). Predicting and interpreting student performance using ensemble models and Shapley additive explanations. *IEEE Access*, 152688–152703.
Salguero, A., Alvarado, C., Griswold, W. G., Porter L. (2021). Understanding Sources of Student Struggle in Early Computer Science Courses. In: *Proceedings of the 17th ACM Conference on International Computing Education Research*, 319–333. https://doi.org/10.1145/3446871. 3469755.
Singh, J., Matthees, B., Odetunde, A. (2021). Learning online education during COVID-19 pandemic – attitudes and perceptions of non-traditional adult learners. *Quality Assurance in Education: An International Perspective*, 29(4), 408-421.
Stewart W.H., Lowenthal, P.R. (2022). Distance education under duress: a case study of exchange students' experience with online learning during the COVID-19 pandemic in the Republic of Korea. *Journal of Research on Technology in Education*, 54, S273-S287.
Suszyński, W., Stęgierski, R., Charytanowicz, M. (2020). Remote access to networking laboratory, secure and scalable solutions. In: *The 2020 International Conference on Computational Intelligence, Information Technology and Systems Research*, 22–23.
Tang, C., Tucker, C., Servin, C., Geissler, M. (2018). Computer Science Curricular Guidance for Associate-Degree Transfer Programs. In: *SIGCSE '18: Proceedings of the 49th ACM Technical Symposium on Computer Science Education*, 435–440. https://10.1145/3159450. 3159536.
Wade, C. (2020). *Hands-On Gradient Boosting with XGBoost and scikit-learn: Perform accessible machine learning and extreme gradient boosting with Python*. Packt Publishing.
Zhang, K and Wu, X. (2022). Synchronous Online Learning During COVID-19: Chinese University EFL Students' Perspectives. *SAGE Open*, 1-10. https://doi.org/ 10.1177/2158244022109482.

M. Charytanowicz is a research scientist at the Department of Computer Science at Lublin University of Technology, as well as at the Systems Research Institute of the Polish Academy of Sciences. The field of her scientific activity to date covers the areas of information technology, computer science education and data mining – mostly connected with the use of modern programming methods and computational intelligence. She has served as a member of scientific committees of international conferences and scientific boards of journals, as well as in the editorship of books.