

# Creating a Metamodel for Predicting Learners' Satisfaction by Utilizing an Educational Information System During COVID-19 Pandemic

Zoe KANETAKI, Constantinos STERGIYOU, Georgios BEKAS, Christos TROUSSAS,  
Cleo SGOUROPOULOU  
*University of West Attica, Athens, Greece*

**Abstract.** Faced with the disruption generated by the COVID-19 pandemic, the advent of enforced and exclusive online learning presented a challenging opportunity for researchers worldwide, to quickly adapt curricula to this new reality and gather electronic data by tracking students' satisfaction after attending online modules. Many researchers have looked into the subject of student satisfaction to discover if there is a link between personal satisfaction and academic achievement. Using a set of data, filtered out of a statistical analysis applied on an online survey, with 129 variables, this study investigates students' satisfaction prediction in a first-semester Mechanical Engineering CAD module combined with the evaluation and the effectiveness of specific curriculum reforms. A hybrid machine learning model that has been created, initially consists of a Generalized Linear Model (GLAR), based on critical variables that have been filtered out after a correlation analysis. Its fitting errors are utilized as an extra predictor, that is used as an input to an artificial neural network. The model has been trained using as a basis the 70% of the population (consisting of 165 observations) to predict the satisfaction of the remaining 30%. After several trials and gradual improvement, the metamodel's architecture is produced. The trained hybrid model's final form had a coefficient of determination equal to 1 ( $R = 1$ ). This indicates that the data fitting method was successful in linking the independent variables with the dependent variable 100 percent of the time (satisfaction prediction).

**Keywords.** machine learning, students' satisfaction, CAD, COVID-19, online learning, hybrid model, Engineering education.

## 1. Introduction

As recognizing the customers' needs is critical to business success, in the educational field, recognizing the learners' needs and taking measures to satisfy them can be the key to enhance students' academic achievements. Many scholars have explored the topic of students' satisfaction in order to discover whether a connection can be established between personal satisfaction and academic performance. It has been described as a comparison of expectations and perceived service quality [1]. However, student satisfaction can be considered as a good indicator of retaining existing students [2], and especially during their first year in higher education. Due to the global COVID-19 pandemic outbreak, traditional teaching methodologies needed to be reformed through

online platforms. Technology features and social media channels have been applied in the field of education in order to create challenging virtual learning environments, especially for the first semester students that have not even visited the University Campus due to lockdown measures.

In view of the above, this paper presents a model that predicts the students' satisfaction in a first-semester mechanical engineering course. The model consists of 24 critical variables of an extended online survey. The novelty of this paper lies in the forecasting of students' satisfaction in learning environments controlled fully by pandemic restrictions.

### *1.1. Related work*

It is important to investigate significant factors in online learning that indicate the success of the method applied. Those success factors can be measured in terms of network education platforms corresponding to the needs of instructors and learners, remote teaching on completing learning tasks efficiently and whether online education can become an efficient tool for specific periods. Outcomes can reveal recommendations based on the research findings, in order to support the sustainability of online education strategies [3]. In [4], the online portion of a blended-learning degree program for pharmacists has been evaluated, using a formal self-assessment and peer review. An instrument systematically devised according to Moore's principles of transactional distance [5] has been applied and the research pointed out that a number of course elements for adjustment could enhance the structure, dialog, and autonomy of the student learning experience. In [6] a virtual reality tour-guiding network has been constructed, and 391 students from a Taiwanese technical university took part in the experiment. The findings of this research revealed their learning efficacy and acceptance of technology in the educational system. During the first stage of the outbreak, in most educational establishments online education took the form of class-based instruction and is an expansion of the original online education.

Previous research on the satisfaction of online education platforms did not consider the new variables introduced by the epidemic, such as ease of use and interaction quality [3]. Nevertheless, new constructs categories have been revealed in [7] such as Students transition from Face-to-Face learning to an Emergency remote Teaching environment [8] and virtual classroom fatigue, that can affect learners' attitude towards a specific online curriculum.

The present study evaluates the learning strategy of an online first-year engineering curriculum from the viewpoint of students, in the context of public health emergency. Factors referenced in previous studies [9] have been optimized, and shown to contribute on student's perception of assignments relating to real world tasks.

## **2. Research methodology**

This research has been conducted on students (population) completing their first semester during the academic year 2020-2021 at the University of West Attica, School of Engineering, Department of Mechanical Engineering (Athens, Greece). The curriculum selected is laboratory course, named "Mechanical design, Computer Aided Design CAD I". The online module's learning strategy has been applied to 216 students

which were divided in 11 online groups in MS Teams platform, assisted by a group of 5 instructors (N=216), with a valid number of 165 participants (n=165).

A large amount of information regarding engineering students and their interactions with their virtual learning environment has been obtained. Data were mined out of two web-based surveys<sup>1</sup>, and additional students' related data. Quizzes, freehand drawings of object views (sketches) and CAD drawings of object views including sectional views were among the weekly assignments. The purpose of the mind-on assignments was to improve students' spatial perception. Specific activities were centred on an existing metallic structure on campus, aiming to establish tasks' relevance to real-world mechanical engineering cases, as well as their importance for students' future employment. [10, 11].

The measuring methods of students' satisfaction of the module can be schematized in the following organogram:



Figure 1. Data mined from different sources.

Following the analysis of the collected data, a matrix with the dimensions 129X165 was created (Figure 1,2), where 165 is the number of students and 129 is the number of variables to be evaluated. A statistical analysis was conducted, which included a correlation analysis (Spearman's rank correlation coefficient) in SPSS v20 to highlight the most significant correlations between all of the ordinal variables. The aforementioned analysis filtered out 24 variables that affect students' evaluation of the module, related to their satisfaction during COVID-19 pandemic circumstances.

The methodology performed is illustrated in Figure 2.

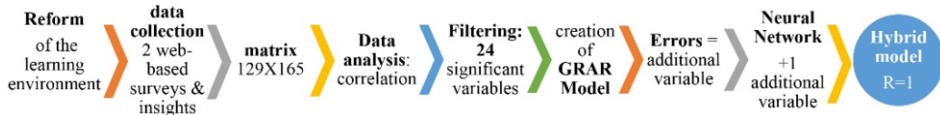


Figure 2. Methodology of the study.

### 3. Metamodel of interpreting students' satisfaction

The presented method aims to combine the benefits of the GLAR method in fitting transformed predictors with linear logic, with the effectiveness of neural networks in non-linear data fitting applications. Neural networks are considered adequate for handling non-linear applications. The machine learning technique known as an artificial neural network (ANN) is used to define a function that connects a set of inputs to a set of outputs. As a result, a generalized linear model generates a linear combination of transformed predictors (through a link function's response).

A separate file containing the most important variables is established for the aim of building a function to associate the students' satisfaction from the module with the database's most influential variables for this pursuit.

<sup>1</sup> 1. pre-course, 2. post-course

The function created will have the form presented below:

$$y = f(x_1, x_2, x_3, x_4 \dots x_n)$$

where  $y$  represents the students' satisfaction and  $x_1, x_2, \dots, x_n$  represent the predictors (table 1), which in this case are the variables (survey questions) that have a statistically significant correlation with the student's satisfaction.

Artificial neural networks are known to be excellent at handling non-linear issues. Therefore, artificial neural networks (ANN) can be conceived as a machine learning methodology for defining the function (metamodel) that connects a set of inputs to a set of outputs. In the relevant literature [12, 13], ANNs have been used in a number of studies.

The procedure mentioned above is performed by connecting the input layer or a set of hidden layers to an intermediate hidden layer (or a set of hidden layers) that is connected to the output layer. The size of the input layer is relative to the number of variables that are being taken into account. Each layer's nodes (also called neurons) pass information on to the nodes that come after them [14-16].

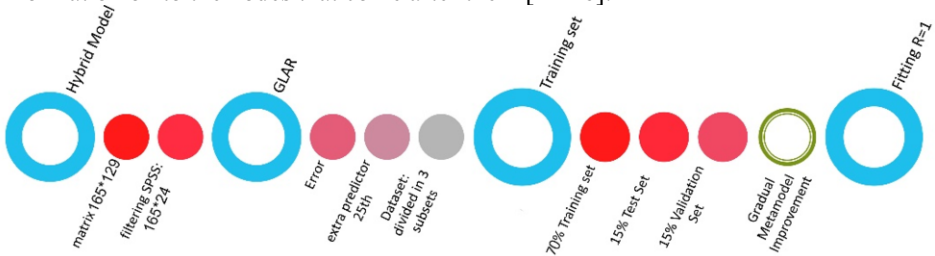


Figure 3. Satisfaction Metamodel neural Network's Architecture.

The suggested method in this study aims to combine the benefits of the GLAR in fitting transformed predictors with linear logic with the efficacy of neural networks in non-linear data fitting applications. The generalized linear model integrates a variety of statistical models, including linear regression, logistic regression, and Poisson regression. It is essentially an iteratively repeatable least squares method that can maximize the likelihood of model parameters.

The survey answers and related data of the 165 students in 24 statistically significant variables (with p-values less than 5% and Spearman rho coefficients equal to absolute 0.20 or more) are isolated in a matrix with the following dimensions: 165x25 (table 1). The students' satisfaction rate is listed in the table's last column (25<sup>th</sup>). In multivariate regressions, the absence of tools dealing with ordinal and nominal variables is a problem that will be handled by the technique applied in the present study. In the process of modelling continuous variables, similar issues do not occur. The errors generated by the Generalized Linear Model will be used in the hybrid model as the 25<sup>th</sup> variable, apart from the 24 variables previously stated. A neural network for predicting student satisfaction from the Mechanical Design CAD I module, is finally trained. All simulations of this research have been performed in MATLAB R2020b.

The process can be described in the following scheme:

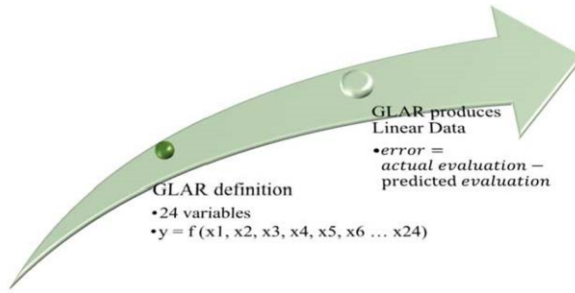


Figure 4. Process scheme.

As a result, the errors originating from the GLAR model are displayed in the histogram (Figure 5).

A normal distribution curve is also included in the histogram in order to compare the generated errors with the ones that would occur if the data were normally distributed (Figure 5). The majority of data are located in the center of the histogram, that represents a range of values between -1 and +1. A general symmetry can be observed with a slight negative left-skewness (-1,024). The left tail of the distribution graph is longer, but since the mean is centrally located on the distribution peak (0,00), it can be concluded that the students' satisfaction prediction variable is normally distributed. As a result, more errors can be found in the region where underestimation of the model's prediction occurs (and not overestimation).

The histogram's bin size was chosen to be equal to 0.5 (Figure 5). It can be seen that the forecasting errors are generally minor.

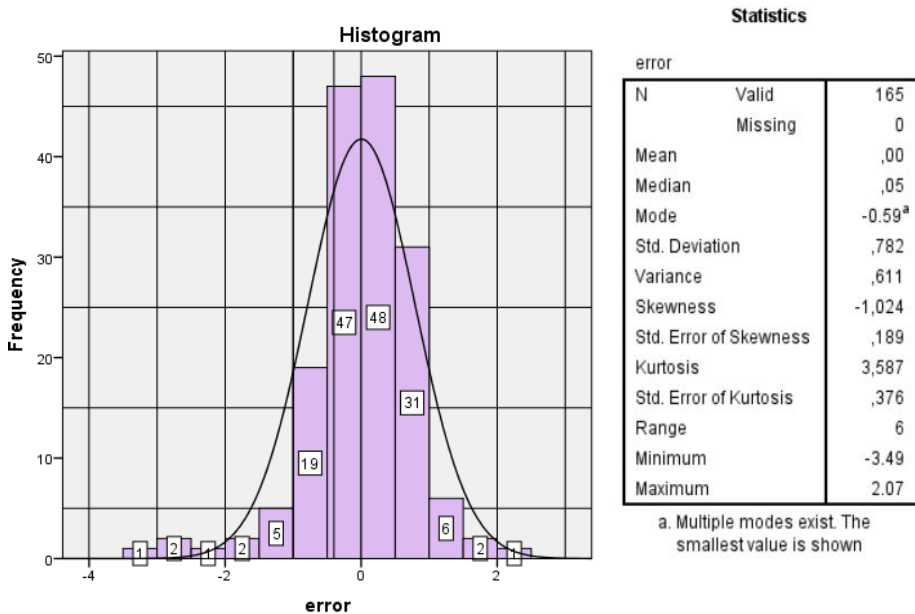


Figure 5. Histogram of Errors and mean error.

The table 1 indicates the values of the coefficient estimates from the GLAR model, which are presented with the following form, which correspond to the equation below:

$$y = \beta_0 + \beta_1x_1 + \beta_2x_2 + \dots + \beta_nx_n + \textit{intercept}$$

where:  $\beta_0, \beta_1, \beta_2 \dots$ , are the coefficient estimates from the GLAR model.

**Table 1.** variables' description, Spearman's RCC and GLAR coefficients.

Variables	Description	Spearman's rank correlation coefficient	coefficients of the GLAR
(Intercept)			0.16
x1	Enjoyable vs other labs	0.527	-0.06
x2	Familiarised to MS Teams vs other modules	0.430	0.06
x3	Satisfied vs other modules	0.673	0.23
x4	Able to complete assignments (difficulties)	0.329	0.00
x5	Satisfied with assignments' grades	0.314	-0.10
x6	Well organised	0.588	0.14
x7	How well tasks are assessed	0.366	-0.32
x8	Theory contributes on accomplishing assignments	0.427	0.14
x9	Evaluate class notes	0.391	0.03
x10	Quizzes contribute on understanding the theory	0.388	-0.03
x11	Evaluate assignments variety (Quizzes, sketches, CAD)	0.451	0.16
x12	Quality of videos in assignments' methodology	0.443	0.06
x13	Enjoyability CAD II lab vs other theoretical modules	0.470	0.28
x14	Classroom fatigue	0.469	0.04
x15	Evaluate interactivity	0.608	0.25
x16	Opportunity to express out loud questions during online lectures	0.418	-0.08
x17	Questions being answered during online lectures	0.485	0.08
x18	Resent late assignment gradings	0.309	0.21
x19	Instructor's comments helped understand mistakes	0.325	-0.05
x20	Assignments related to future work	0.340	0.06
x21	Presentation and clarity of instructions in the 15th assignment	0.376	0.09
x22	Assignments related to Real World Tasks	0.343	0.05
x23	Sustainability of the learning strategy in Face-to-Face learning	0.284	-0.03
x24	Evaluate CAD I vs other modules	0.630	0.31

The majority of errors in 145 instances (out of the total 165 instances) fall into a range between -1 and +1 (Figure 5). Therefore in 87,88% of cases, the errors fall into the aforementioned range.

The following diagrams are obtained: R (coefficient of determination/ prediction success rate; in this case equal to 1), error histogram (with classification of errors per subset (there are three distinct subsets in the present study: training subset (representing 70% of the observations), validation subset (representing 15% of the observations), and test subset (representing 15% of the observations)). The overall artificial neural network performance is also displayed in Figure 6.

It's worth noting that the process of breaking down observations into sets (sets) contributes to the neural network's statistical independence, whereas a model derived from a large set of observations (training set) is used to generate predictions in other subgroups of observations.

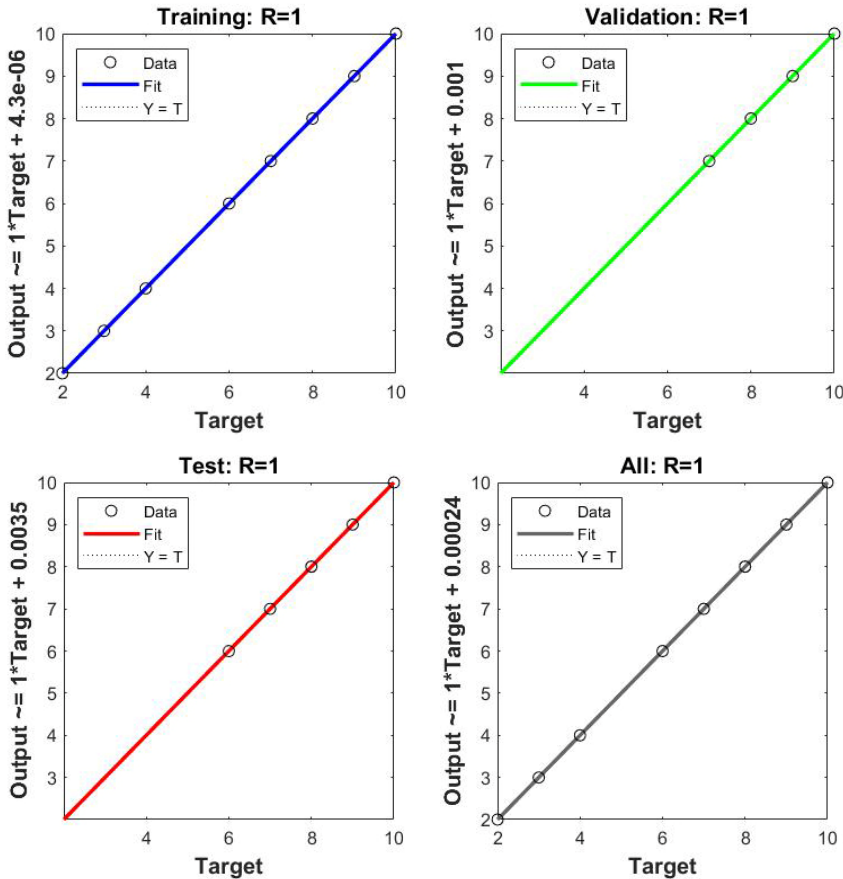


Figure 6. Training set, validation set.

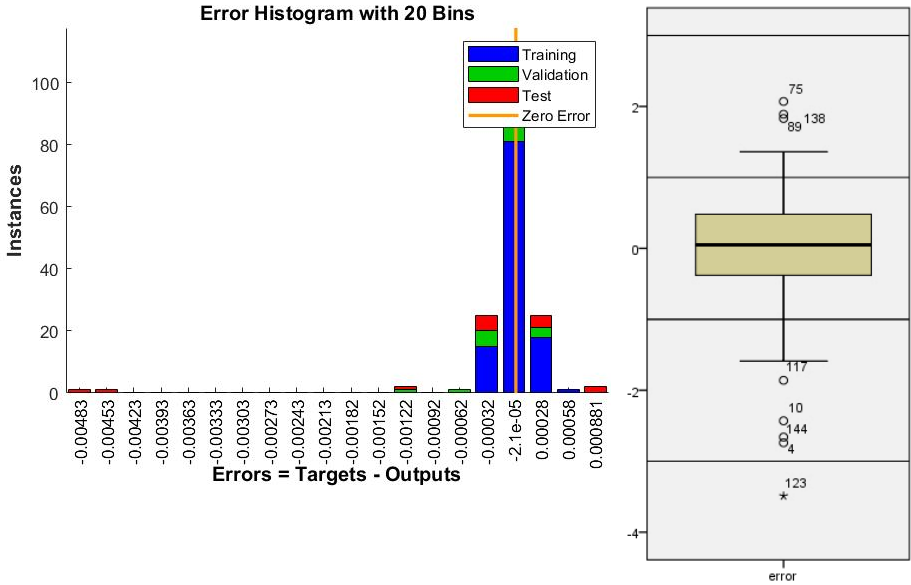


Figure 7. Error Histogram and boxplot.

As shown in the boxplot above (Figure 7), the median is equal to zero. The upper and lower whisker display the position of the first (Q1) and the third (Q3) quartile and predicted values (interquartile range) are mostly gathered between the whiskers, while outliers are displayed above and below the whiskers.

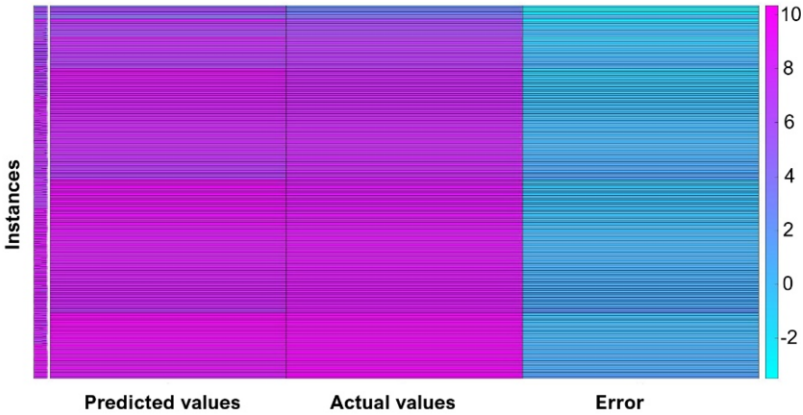


Figure 8. Heatmap of errors

In order to visualize the relation between the actual values and the predicted values a heatmap has been created (Figure 8), where numerical values have been replaced by colors, with three data columns: Predicted values, Actual values and Error. In Figure 8, the color of the cell represents the values. The color gradient in the predicted values is uniform and almost identical to the actual values color gradient. Values in between are shown faded and centered above the average.



#### 4. Conclusions and Future work

It can be concluded that students' satisfaction prediction from a university module attendance during COVID-19 pandemic has not been widely researched until today, whereas satisfaction prediction of online attendance can be a useful component of academic experience [17-21], or even a recommendation tool for selecting specific subjects and university curricula. Students can improve their happiness by their own choices, which represents a great challenge when online learning becomes exclusive and imposed [22, 23]. Recognizing learners' needs and taking steps to meet them might help students achieve greater academic success.

For this purpose, a forecasting tool has been created, by filtering out the critical variables of an extended online survey, applied to students attending a first-semester Mechanical Engineering module, Mechanical Design CAD I, under pandemic circumstances. Therefore, a hybrid model has been created, using 24 critical variables as predictors, focusing on forecasting students' satisfaction. Since there has been no previously similar learning environment controlled by pandemic restrictions, the actual research has been centered on forecasting students' satisfaction.

The final hybrid model was separated into three subsets (Figure 3), with a training set of 70% of the data predicting the test set (15%) and the validation set (15%), resulting in a  $R=1$  fit.

Future work consists of testing the performance of those 24 variables to the academic year 2021-2022 first-semester students, attending the specific module in a similar or blended learning environment in order to determine the overall success of the methodology suggested by this study.

#### References

- [1] R.L. Oliver, A cognitive model of the antecedents and consequences of satisfaction decisions, *Journal of Marketing Research* 17 (1980), 460-469.
- [2] I. Çavar, L. Bulian, N. Dubreta, Student Satisfaction in Higher Education: Factors Affecting Engineering Students' Satisfaction, ICERI2019 Proceedings (2019), 6782-6792.
- [3] T. Chen, L. Peng, X. Yin, J. Rong, J. Yang, G. Cong. Analysis of User Satisfaction with Online Education Platforms in China during the COVID-19 Pandemic. *Healthcare*, (2020).
- [4] K. Wilbur, Evaluating the online platform of a blended-learning pharmacist continuing education degree program. *Med. Educ. Online* (2016).
- [5] M. Moore, *The Theory of Transactional Distance* (2013).
- [6] H.M. Chiao, Y.L. Chen, W.H. Huang, Examining the usability of an online virtual tour-guiding platform for cultural tourism education. *J. Hosp. Leis. Sport Tour. Educ.* (2018) 32, 29-38.
- [7] Z. Kanetaki, C. Stergiou, G. Bekas and E. Kanetaki, "Machine Learning and Statistical Analysis applied on Mechanical Engineering CAD course: A Case Study During ERTE Pahse in the Context of Higher Education," 2020 4th International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT), 2020, pp. 1-13, doi: 10.1109/ISMSIT50672.2020.9254924.
- [8] C. Hodges, S. Moore, B. Lockee, T. Trust, A. Bond, The difference between emergency remote teaching and online learning. *Educause Review*, (2020).
- [9] M. Binkley, O. Erstad, J. Herman, S. Raizen, M. Ripley, M. Miller-Ricci, M. Rumble, *Defining Twenty-First Century Skills*. (2011)
- [10] U. Dominguez, J. Magdaleno, *Active Learning in Mechanical Engineering Education in Spain*. Conference: WEE 2011At: Lisbon, Portugal (2011).
- [11] Z. Tordai, I. Holik, Student's Characteristics as a Basis for Competency Development in Engineering Informatics Education. *International Journal of Engineering Pedagogy*, (2018) 8: 32-42.
- [12] C. Troussas, F. Giannakas, C. Sgourpoulou, I. Voyiatzis, Collaborative activities recommendation based on students' collaborative learning styles using ANN and WSM, *Interactive Learning Environments* (2020). DOI: 10.1080/10494820.2020.1761835.

- [13] F. Giannakas, C. Troussas, I. Voyiatzis, C. Sgouropoulou, A deep learning classification framework for early prediction of team-based academic performance, *Applied Soft Computing* **106** (2021), 107355. <https://doi.org/10.1016/j.asoc.2021.107355>
- [14] K.G. Sheela, S.N. Deepa, Review on Methods to Fix Number of Hidden Neurons in Neural Networks. *Mathematical Problems in Engineering*, Hindawi Publishing Corporation (2013).
- [15] P. Yao, Integrating Generalized Linear Auto-Regression and Artificial Neural Networks for Coal Demand Forecasting. *Advances in Neural Networks - ISNN 2009* (2009) pp. 993-1001.2009.
- [16] M.W. Trosset, *An Introduction to Statistical Inference and Its Applications with R*; CRC Press, Taylor & Francis Group: Boca Raton, FL, USA (2009).
- [17] Z. Mihanović, A. Batinić, J. Pavičić, The link between students' satisfaction with faculty, overall students' satisfaction with student life and student performances. *Review of Innovation and Competitiveness*. (2016) 2. 37-60. 10.32728/ric.2016.21/3.
- [18] I. Duzevic, Student satisfaction vs student achievements – should quality management system in higher education aim at student satisfaction or student achievements? *Poslovna izvrsnost - Business excellence*. (2020) 14. 51-67. 10.22598/pi-be/2020.14.2.51.2020
- [19] C. Papakostas, C. Troussas, A. Krouska, C. Sgouropoulou, User acceptance of augmented reality welding simulator in engineering training. *Education and Information Technologies* (2021). <https://doi.org/10.1007/s10639-020-10418-7>
- [20] C. Troussas, A. Krouska, C. Sgouropoulou, Impact of social networking for advancing learners' knowledge in E-learning environments. *Education and Information Technologies* (2021). <https://doi.org/10.1007/s10639-021-10483-6>
- [21] C. Papakostas, C. Troussas, A. Krouska, C. Sgouropoulou Measuring User Experience, Usability and Interactivity of a Personalized Mobile Augmented Reality Training System. *Sensors* 21(11) (2021), 3888. <https://doi.org/10.3390/s21113888>.
- [22] M. Omar, Influence of Wi-Fi Performance on Students' Satisfaction. *International Journal of Modern Education*. 3. 226-237. 10.35631/IJMOE.380018.2021
- [23] Kéri, Anita. Online Teaching Methods and Student Satisfaction during a Pandemic. *Language Learning* (2021)