MLS - EDUCATIONAL RESEARCH

http://mlsjournals.com/Educational-Research-Journal

ISSN: 2603-5820

How to cite this article:

Domínguez Azpíroz, I., García Villar, M., Brito Ballester, J., Rodríguez Velasco, C. L. & Soriano Flores, E. (2023). Clasificación y pronóstico del nivel de satisfacción de egresados de programas de salud en el contexto de una metodología de aprendizaje automático: un análisis de caso orientado a posgrados online de una institución educativa iberoamericana. *MLS-Educational Research, 7*(2), 88-115. 10.29314/mlser.v7i2.1721.

CLASSIFICATION AND FORECAST OF THE LEVEL OF SATISFACTION OF GRADUATES FROM HEALTH PROGRAMS IN THE CONTEXT OF A MACHINE LEARNING METHODOLOGY: A CASE ANALYSIS ORIENTED TO ONLINE POSTGRADUATE DEGREES FROM A LATIN AMERICAN EDUCATIONAL INSTITUTION

Irma Domínguez Azpíroz
European University of the Atlantic (Spain)
irma.dominguez@uneatlantico.es - https://orcid.org/0000-0001-7261-5205
Mónica Gracia Villar
European University of the Atlantic (Spain)
monica.gracia@uneatlantico.es - https://orcid.org/0000-0002-8547-9246
Julién Brito Ballester
European University of the Atlantic (Spain)
julien.brito@uneatlantico.es - https://orcid.org/0000-0001-6436-0214
Carmen Lili Rodríguez Velasco
European University of the Atlantic (Spain)
carmen.rodriguez@uneatlantico.es - https://orcid.org/0000-0002-9609-4026
Emmanuel Soriano Flores
European University of the Atlantic (Spain)
emmanuel.soriano@uneatlantico.es - https://orcid.org/0000-0002-8747-5679

Abstract. The purpose of this research article was to perform a classification based on neural networks to predict the level of satisfaction of a sample of graduates, corresponding to different graduate programs in the health area of a Latin American educational institution under an e-learning methodology. To this end, a Likert scale questionnaire model was instrumented which, after being validated, had a reliability of 0.791. Likewise, the average global satisfaction index of the graduates was 2.66/4, with a better score in the section on logistics of materials and in the management and technical support of the virtual campus, while the lowest scores referred to aspects related to extra-center communication and the facilities offered by the institution for the improvement of the participant's economic and social context. Finally, the probabilistic classification and prediction algorithm of the neural network obtained an accuracy of 96.8%, indicating an excellent degree of model fit. The methodology followed and the rigor in determining the validity and reliability of the instrument, as well as the subsequent analysis of the results, endorsed by the review of the documented information, suggest that the instrument can be applied to other multidisciplinary programs for decision making with guarantees in the educational field.

Keywords: health, graduate satisfaction, Likert scale, neural network, postgraduate program

CLASIFICACIÓN Y PRONÓSTICO DEL NIVEL DE SATISFACCIÓN DE EGRESADOS DE PROGRAMAS DE SALUD EN EL CONTEXTO DE UNA METODOLOGÍA DE APRENDIZAJE AUTOMÁTICO: UN ANÁLISIS DE CASO ORIENTADO A POSGRADOS ONLINE DE UNA INSTITUCIÓN EDUCATIVA IBEROAMERICANA

Resumen. El propósito de este artículo de investigación fue realizar una clasificación basada en redes neuronales, para pronosticar el nivel de satisfacción de una muestra de egresados, correspondiente a diferentes programas de posgrado del área de salud de una institución educativa latinoamericana bajo una metodología e-learning. Con este fin, se instrumentalizó un modelo en un cuestionario de escala de Likert que, tras ser validado, resultó con una confiabilidad de 0.791. Asimismo, el índice global medio de satisfacción de los egresados fue de 2.66/4, observando una mejor puntuación en el apartado de logística de materiales y en el manejo y soporte técnico del campus virtual, mientras que las puntuaciones más bajas se refirieron a aspectos relacionados con la comunicación extra-centro y las facilidades ofrecidas por la institución para la mejora del contexto económico y social del participante. Finalmente, el algoritmo de clasificación y predicción probabilística de la red neuronal obtuvo una precisión del 96.8%, lo que indicó un excelente grado de ajuste del modelo. La metodología seguida y el rigor en la determinación de la validez y confiabilidad del instrumento, así como el posterior análisis de resultados, refrendado con la revisión de la información documentada, hace presuponer la aplicación del instrumento a otros programas multidisciplinares para la toma de decisiones con garantías en el ámbito educativo.

Palabras clave: salud, satisfacción de egresados, escala de Likert, red neuronal, posgrados

Introduction

The changing paradigm of distance education in the health field

Distance education - in its non-school and blended modalities - represents a paradigm of the teaching-learning concept, alternative to the traditional method, which allows reaching a greater number of people in a global scenario based on knowledge.

Implicit in this paradigm shift is the incorporation of virtual university models that offer students the advantage of being able to choose a time and place for learning. This is important in the field of postgraduate continuing health education, especially in those healthcare groups located in rural or inaccessible environments (Domínguez, 2021).

In this sense, research has focused more on establishing relationships between different variables than on specifically defining the profile of the graduate of a distance program (Pérez, Martínez, & Martínez, 2015; Álvarez, Chaparro, & Reyes, 2015; Surdez, Sandoval, & Lamoyi, 2018), hence the need for studies that serve to:

> [...] continue research along the lines of a more participant-centered model, implemented on reusable software platforms for specific learning contexts, with the purpose of offering "tailor-made" training for the intended user, especially in unexplored contexts (Fainholc, 2016).

Therefore, current trends in e-learning in health education will require a commitment on the part of educators to use technologies that facilitate this curricular shift toward learner self-regulated learning (Brydges et al., 2015).

However, it was not until a few years ago, with the advent of information and communication technologies (ICT), that a series of theories and models of e-learning training evaluation emerged with partial and global approaches, under different perspectives.

For example, the Knowlege, Process, Practice (KPP) e-learning training model, proposed by Shaw, Barnet, McGregor, and Avery (2015), for its application to different professional groups in the health area, proved to be flexible to technological progress by guaranteeing not only the delivery of knowledge, but also the way to process it through satisfaction assessment tools and application in practice.

In parallel, the World Health Organization (WHO, 2005) called on all member governments to adopt and use information technologies for the benefit of public health, through Resolution WHA58.28 eHealth, and in particular, to promote equitable enjoyment at an affordable price, reduce the digital divide and:

[to continue the extension to Member States of mechanisms such as the Health Academy, which promote healthy lifestyles and a better understanding of health-related issues through e-learning (p.128).

In this context, the term *e-Health*, also referred to as digital health, is a comprehensive concept that brings together a group of applications, associated with ICTs, that have helped to strengthen, advance and create opportunities in the performance of health-related fields (Shiferaw and Zolfo, 2012; WHO, 2018).

Despite this, the heterogeneous nature of the models and the lack of concreteness of the concept of quality in the context of evaluation in general, have prevented the definition of universal criteria for assessing student satisfaction (Pereira and Gelvez, 2018).

Models for measuring the level of student satisfaction in virtual training

The relationship between the satisfaction of graduates and the degree of educational quality constitutes one of the issues that has most concerned Universities, interested for years in improving their offer to attract new generations of students to the different higher education modalities (González, Tinoco and Torres, 2017).

In this context, knowing the level of satisfaction of postgraduate graduates can help educational institutions to: attract new generations of students (Gónzalez et al., 2016), respond to [national] and international university evaluation bodies (Pérez et al., 2015), find out their needs

(Mejías and Martínez, 2013) and, finally, obtain a better positioning in academic performance among higher education institutions (Surdez et al., 2018).

Consequently, it is important to contemplate quality management models and standards that indicate as a requirement the need to establish a process for measuring customer or user satisfaction (Pérez et al., 2015). However, compared to face-to-face education, there are few models aimed at e-learning training and even fewer that consider graduate satisfaction as an element to be taken into account in decision-making.

One of these few tools is the standard "Quality Management. Quality of Virtual Training (UNE 66181:2012)", which determines the three main factors involved in meeting the needs and expectations of students: recognition of training for employability, learning methodology and accessibility.

Table 1 shows a definition of the factors that have been considered most important during the development of this research.

Table 1

Factors that condition student satisfaction in virtual training

Factors	Description
Accessibility	Ability of the training program to meet the student's accessibility needs and expectations due to digital divide issues, disabilities.
Learning methodology	Ability for the training program to efficiently find solutions to complex problems, based on analysis and in the context of methodological planning
Cost	Monetary cost of the training program
Maintainability	Potential ability for the training program to be modified, expanded, etc.
Employability	Ability of the training program to facilitate entry into the labor market

Note: Adapted from UNE 66181:2012

Instrumentalization of the model

The measurement instrument must contain the measurement criteria and have a minimum level of abstraction, so it is necessary that the variables of the model undergo a process called "operationalization of variables", which moves from the general to the particular, from the abstract to the concrete.

On the other hand, the instrument must be valid and reliable. In this context, validity can be content, criterion and construct validity. In general, it is common for models to carry out content validity by means of a literature review and consultation with panels of experts (Mejías and Martínez, 2013). In relation to how reliable the model is, the reference indicator is the "Cronbach's Alpha" statistic, widely used in the university environment (González and Pazmiño, 2015), and where the approximation to unity is the most desirable.

Artificial intelligence: Machine Learning and Deep Learning

Artificial intelligence is an important tool to help close the gap between graduate satisfaction and educational quality, contributing to the paradigm shift that virtual education represents.

According to Pedemonte (2020), Artificial Intelligence (AI) is defined as a software technology that encompasses one or more capabilities related to perception, prediction, classification, decision making, diagnosis and logical reasoning, among others.

For its part, machine learning or Machine Learning refers to the ability of the machine to learn on its own based on experience, from the interpretation of a set of inputs, and thus provide one or more outputs to meet a given objective (Faggella, 2018).

The machine learning methodology has been used to develop what is now known as "Deep Learning", a machine learning system based on a sequential set of layers of artificial neural networks, which attempt to mimic the functioning of neurons in the human brain (Schmidhuber, 2015).

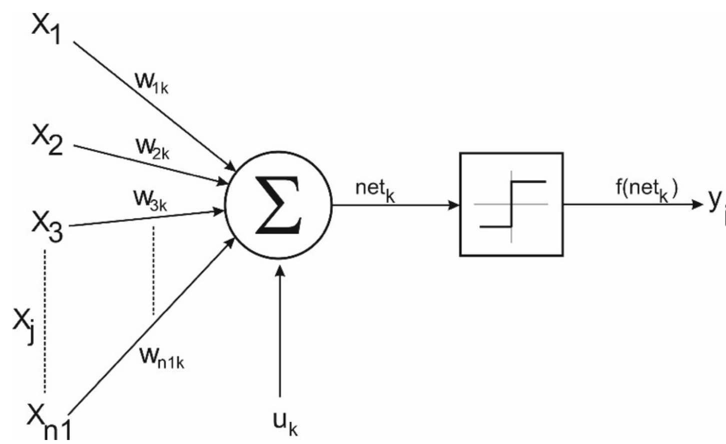
Simple architecture of the artificial neural network

In practice, the materialization of this methodology translates into a mathematical model consisting of a series of connected processing units or neurons, which attempt to mimic the functioning of the biological neural network of the human brain.

The simplest model of an artificial neural network (ANN), also called a perceptron, consists of an input layer, a neuron and a single output, whose connections are associated with weights (w_{jk}), which give an idea of the intensity of the input signal. Each neuron also has a threshold or bias (u_k) associated with it, which functions as a switch to turn the neuron on or off (Figure 1).

Figure 1

Simple perceptron model



In neural network models, a linear combination of the weights multiplied by the inputs is performed to subsequently determine the output based on the resulting sum, by means of a continuous, differentiable and nonlinear activation function (sigmoid, ReLu, hyperbolic tangent...), since the aim is to obtain values as close as possible to 0 and 1 at the output (feed-forward):

$$net_k = \sum_{j=1}^{n1} w_{jk} \cdot x_j + u_k \qquad (1)$$

and the exit:

$$y_i = f(net_k) \qquad (2)$$

where f is the activation or transfer function.

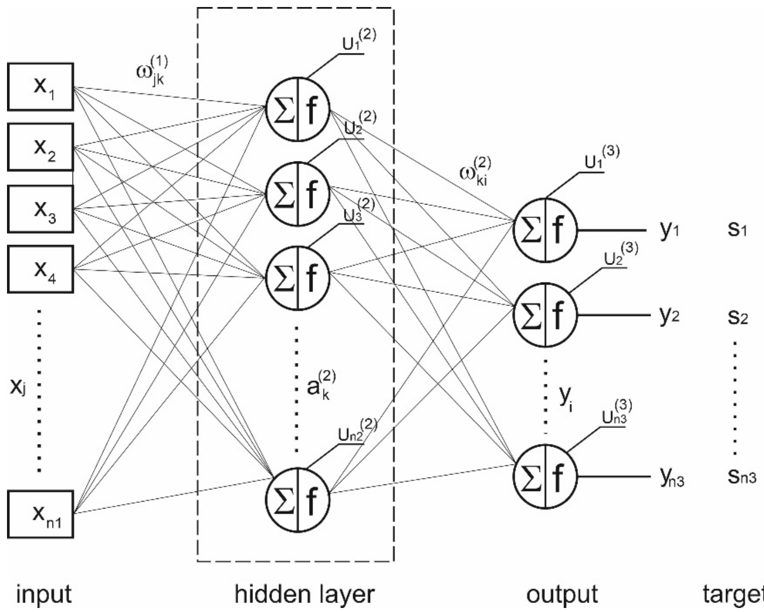
Architecture of a multilayer neural network

However, the applications of the simple perceptron are very limited, so it is common to add layers of neurons (called hidden) between the input and output, which adds complexity to the network.

In this sense, Figure 2 illustrates the architecture of a multilayer neural network composed of an input layer, a hidden layer and, finally, an output layer.

Figure 2

Architecture of a multilayer neural network. Note: Activation functions can be different for each layer



The choice of the number of hidden layers and their composition conditions the efficiency of learning and influences the generalization of the network (Castillo, Solórzano, & Moreno, 2018). Normally, in Machine Learning methodology, unlike Deep Learning, a single hidden layer is sufficient for the algorithm to converge, although its composition is determined by trial-and-error testing. However, as a guideline, the number of neurons in the hidden layer is usually chosen to be equal to the average of the number of neurons in the input and output layers.

Train, Validation and Testing

During network learning, overfitting problems may occur, where the network will perform excellently for the training patterns, but will have poor generalization capability, i.e., it will not be able to respond adequately for values outside the training.

In order to avoid overfitting, the input data set is divided into three parts: train (train+validation) and testing, so that for each epoch a training error and a validation error are provided. Obtaining the minimum validation error is the signal to finally exit the algorithm, avoid overfitting and start the test stage.

Testing data is used to test the model with new inputs that have not been trained and validated.

Research design

In previous sections, the importance of having a valid and reliable instrument to measure the level of satisfaction of the graduates of various online graduate programs in the field of health was highlighted. However, the heterogeneity of norms and standards has so far prevented the definition of universal measurement criteria for their application in practice.

Thus, once the problem was posed, the research question was as follows:

- is it possible to develop a deep learning methodology, which has as input parameters a set of measurement criteria, to classify and make predictions about the level of satisfaction of graduates of online graduate programs in the health area?

Table 2 shows the guidelines for this research.

Table 2
Research guidelines

Unit of analysis:	Satisfaction of graduates of online postgraduate programs in the health field
Dependent or output variable:	Level of satisfaction of graduates of online postgraduate programs in the health field
Operational definition of the variable	⇩
Values of the dependent variable:	high, medium, low
Independent or input variables:	Measurement criteria
how is the data collected?	⇩
Unit of observation:	Set of Likert-scale responses from graduates administered by panels of experts

Note: Adapted from Martinsuo & Huemann (2021) and Azcona & Manzini (2019)

In this context, the research sub-questions were as follows:

- Why is it necessary to operationalize the variables of a model?
- Which measurement criteria are most valued by graduates? Which ones need to be improved?
- Can artificial intelligence reduce the existing gap between the level of student satisfaction and the degree of educational quality in health postgraduate programs?

- Is it possible to determine the probability of predicting the level of satisfaction of a graduate health graduate from the answers given to the measurement criteria of a questionnaire or instrument?

Method

The methodology followed in this research was descriptive and correlational, with a quantitative, non-experimental, transectional approach, since no hypotheses were posed and no variables were manipulated, but "[...] data were measured, evaluated or collected on various aspects, dimensions or components of the phenomenon to be investigated [in its natural work environment and in a single time]" (Hernández, Fernández & Baptista, 2003; Pérez et al., 2015).

The diagnosis of the level of satisfaction of graduates with the postgraduate programs in question consisted of different phases: first, a panel of experts from the University elaborated an abstract model using theories, models and evaluation tools for virtual training; second, after a process of operationalization of the variables, a set of 13 measurement criteria was obtained, which formed part of a satisfaction questionnaire on a Likert scale, which was applied, validated and its reliability determined on a total of 241 participants; finally, based on the measures of central tendency and dispersion, two decision thresholds were found to classify the results obtained into a high, medium or low level of satisfaction. This last result was used to train, validate and test a neural network, and to establish a forecast of the level of satisfaction of new students using Matlab R2021b® mathematical software.

The panel of experts was made up of two doctors in the area of health and nutrition, a professor in information technology and a doctor in projects, who previously established common guidelines to reach a good level of consensus.

Main theories, models and tools for e-learning

Table 3 shows the main bibliographic references used to determine the variables, dimensions, factors and indicators of the model.

As shown, the compilation of bibliography was oriented towards training activities and logistical and technological infrastructure.

Table 3

Main virtual education models and associated variables

Sphere of influence	Models	Variables
Training action	Systemic model of Vann Slyke et al. (1998)	– Institution implementing the training action – Capacity of training recipients – Adaptability of the *e-learning* system – Adaptation of users to the virtual campus
	Marshall & Shriver's (1994) five-level model, in McArdle (1999)	– Communication skills of the teacher – Course materials (difficulty, relevance...) – Content or Curriculum – Modular structure – Transfer of learning
	Kirkpatrik's four-level model (1994)	- Participant satisfaction - Assessment of learning, behavior and results
	Marcelo & Zapata Model (2008)	– Socio-cultural context of the participant – Design of objectives and strategies – Facilitation of resources – Virtual learning environment – Tutoring – Continuous evaluation – Follow-up
	Marciniak's (2015) virtual education evaluation model. *Benchmarking* Methodology	– Institution's strategic plan – Institutional context – Educational agents – Teaching-learning processes – Virtual platform
Logistics and technological infrastructure	SCORM	- Content aggregation - Execution environment - Sequencing and navigation
	IMS *(Global Learning Consortium)*	- Metadata - Sequencing - Virtual content design

Design of the instrument based on the measurement criteria

In its most abstract conception, the model investigated considered the following variables: "Gender", "Origin", "Program", "Age", "Entry Profile" and, finally, "Graduate Satisfaction".

The high degree of abstraction of the variable "Graduate Satisfaction" made it necessary to transform it, through a process called operationalization of variables, in order to make it more observable and measurable (Reguant and Martinez, 2014). In this way, five dimensions were considered to measure the degree of satisfaction of the graduates of the *online* postgraduates in question (Figure 3).

Figure 3

Dimensions of the variable "Graduate Satisfaction"

Although the degree of abstraction of the dimensions allows at this time to propose a graphic visualization, the same is not true if later on the variable has to be instrumentalized in a Likert scale questionnaire, so it is necessary to continue to submit the dimensions to a higher degree of concretization, progressively going from the general to the particular, based on the following sequential stages: factors, indicators and measurement criteria (Reguant and Martinez, 2014).

Once the list of factors and indicators was available, the measurement criteria for the diagnosis were developed through a bibliographic search and the contribution of a panel of experts.

In this way, 13 items or measurement criteria were obtained, which provided the basis for a Likert scale questionnaire, with categories "1. Strongly disagree"; "2. Disagree"; "3. Agreed" and "4. Totally agree", to measure the variable of graduate satisfaction with the reference postgraduate programs.

Population and Sample

Initially, the target population was a total of 325 graduates of postgraduate programs in the area of health at the Universidad Europea del Atlántico. In order to determine the necessary sample size, and given that the intention was to estimate percentage distributions of qualitative variables in the statistical calculations, the following formula for finite populations was used (Torres and Karim, 2021):

$$n \geq \frac{N * Z^2_{1-\frac{\alpha}{2}} * (p * q)}{(N-1) * \varepsilon^2 + Z^2_{1-\frac{\alpha}{2}} * (p * q)} \qquad (3)$$

where:

- n = required sample size.
- N = population size.
- $Z_{1-\alpha/2}$ = 1.96 (Z-statistic, calculated at 95% confidence level).
- $p = q$ = 0.5 (typical values under worst-case conditions).
- Error (ε) = 0.05.

The sampling was convenience sampling, i.e., non-probability.

Substituting the values in the formula resulted in a required sample size for the study of $n \geq 176$

General characteristics of graduates

Once the Likert scale questionnaire was applied to 241 participants, a total of 54 missing values were reported in the items (22.4%), due to gaps in the answers which, added to an omission described in the "Origin" variable, totaled 55 cases, which were extracted from the analysis, so that the number of valid graduates included for the diagnosis was finally 186.

Table 4 shows the definitive general characteristics of the graduates, once the data for the analysis had been cleaned.

Table 4

General characteristics of graduate program graduates (N=186)

Nominal variables	Category	n	%	Ordinal variables	Category	n	%
Genre	Male	56	30.1	Age group (years)	20-29	50	27
	Female	130	69.9		30-39	70	37
Source	North America	54	29.2		40-49	39	21
	Central America	47	25.2		50-59	21	12
	South America	72	38.5		60-69	6	3
	Eurasia	13	7.2	Entry Profile	PhD	8	4.4
Program	Endocrinology and nutrition	20	10.6		Master's Degree	30	16.0
	Preparation of diets	13	6.8		Postgraduate	23	12.3
	Clinical psychology and nutrition	7	3.8		Grade/Dip/ Lic	125	67.3
	Physical activity and nutrition	31	16.6				
	Culinary techniques	31	16.6				
	Health and nutrition	85	45.6				

It was observed that 69.9% were female and the remaining 30.1% were male; 38.5% came from South America, 29.2% from North America, 25.2% from Central America and 7.2% from Eurasia. In relation to age, the 30-39 age group accounted for 37%, the 20-29 age group for 27%, the 40-49 age group for 21%, the 50-59 age group for 12% and the 60-69 age group for 3%. In reference to previous studies, 67.3% have completed a degree/diploma/licensure, 16% a Master's degree, 12.3% a postgraduate degree and 4.4% a doctorate. Finally, 45.6% have taken the Health and Nutrition program, 16.6% the Culinary Techniques program, another 16.6% the Physical Activity and Nutrition program, 10.6% the Endocrinology and Nutrition program, 6.8% the Diet Development program and 3.8% the Clinical Psychology and Nutrition program.

Variable "Level of graduate satisfaction"

The variable "Level of satisfaction" was created as a result of the sum of the items for each of the 186 participants, and was the one taken from now on as a reference for the study. This new column (dependent variable or target) included the categorization of the corresponding level of satisfaction (low, medium, high) for each graduate.

Artificial neural network design and implementation

Spreadsheet data preparation

In this first stage, a matrix (186 x 13) was created in the Excel v2016® program, corresponding to the input set of the neural network, which contained the evaluations of the graduates in relation to the 13 questions or items.

For the purposes of its implementation in the algorithm, this column was coded into dummy variables (Table 5).

Table 5

Coding of satisfaction level categories in dummy variables

Level of satisfaction	Target coding
Under	0 0 1
Medium	0 1 0
High	1 0 0

Matlab R2021b® mathematical software matrix import

The next step consisted of importing the described matrix into Matlab R2021b® mathematical software and storing the input and target data in two different matrices.

Division of data

The data were divided into three sets: training, validation and testing.

- Seventy percent of the data was used for network training, i.e., for gradient calculation and updating of network weights and biases.
- Fifteen percent of the data was used for network validation. These data, which are also part of the training, were used to find the best (most likely) model and stop the training to avoid overfitting the neural network.
- The remaining 15% of the data were used to test the optimal generalization of the network using data not used during training.
-

Data normalization or scaling

Data normalization or scaling was performed according to equation 4 in the interval [-1 1]:

$$y = \frac{(y_{max} - y_{min}) \cdot (x - x_{min})}{(x_{max} - x_{min})} + y_{min} \qquad (4)$$

Creation of the neural network model architecture

Seventeen neurons were added in the input layer, corresponding to the assessments of each of the SDGs, another 10 in the hidden layer and, finally, 3 in the output layer.

The hidden layer activation and output functions were of type tansig and softmax, respectively, as these functions will provide the probability that the project has a high, medium or low level of sustainability.

Neural network training and validation

Once the data had been prepared and the architecture designed, the neural network training was performed.

The Matlab R2021b ® function "trainscg" was chosen for the update of weights and biases, according to the gradient descent method and, for the calculation of losses between predictions and target data, the function "crossentropy".

Parallel to training, Matlab R2021b ® performed the validation of the corresponding data set for each iteration, determining an average minimum validation error, which served to delimit the most probable best model, stop training and avoid overfitting, after confirming a decrease in the training error curve together with an increase in the validation error, consecutively, during six more iterations (early stopping technique).

Model testing

Once the network was trained, during the testing phase, data not used during training were introduced in order to check the optimal generalization of the model.

Results

Selection of nominal and ordinal qualitative variables for the model

The model was composed of nominal and ordinal qualitative variables. Among the first, "Gender", "Origin" and "Program" studied by the graduate were considered. In reference to the ordinal qualitative variables, "Age", "Entry Profile" and "Graduate Satisfaction" were identified.

Variables, dimensions, factors and indicators

Table 6 shows the indicators identified for the variable "Graduate Satisfaction", which will be used as measurement criteria for the development of the instrument.

Table 6

Variables, dimensions, factors and indicators identified for the "graduate satisfaction" variable

Variable	Dimensions	Factors	Indicators
Graduate satisfaction	Training	Initial expectations	Degree of satisfaction achieved
		Relevance for training	Degree of learning achieved
		Difficulty of the academic program	Level of difficulty of the academic program
	Participant's context	Economic status	No. of scholarships granted, payment facilities.
	Communication	Interaction between participants, tutors and other interested parties	Degree of participant satisfaction with the channels established for external communication
		Product and/or service information	Level of transparency and truthful information about the academic program
	Teaching-learning methodology	Academic Program	Academic program subjects
		Evaluation	Continuous evaluation activities
		Teacher evaluation	Degree of participant satisfaction with academic tutors
			Degree of participant satisfaction with the PFM mentor
	Resources	Accessibility to the product and other services offered by the institution	Number of shipments of teaching material, reception times, etc
		Virtual Campus	Ease of use of the virtual campus
			Technical support and number of incidents

Validation and reliability of the measuring instrument

Table 7 shows the measurement instrument adapted to a Likert-type scale.

Table 7

Measurement instrument or Likert scale questionnaire for measuring graduate satisfaction with the graduate program

Dimension	Item No	Measurement criteria
Training	1	The academic program has met my initial expectations
	2	The tutor has made this program relevant to my training and professional performance
	3	The degree of difficulty of the program has been higher than others I have taken
Participant's context	4	The Institution has granted me economic facilities to be able to carry out the study
Communication	5	I am satisfied with the attention received prior to enrollment
	6	I found the academic information provided during the program to be sufficient
Teaching-learning methodology	7	The curricular structure of the program seemed to me to be very appropriate
	8	My assessment of the continuous evaluation is very satisfactory
	9	My mentor has gone out of his/her way to help me
	10	My Master's Final Project director has been accessible
Resources	11	Delivery of didactic materials has been punctual and on time
	12	The handling of the virtual campus has been very user friendly
	13	My assessment of the technical support service of the virtual campus is very satisfactory

The validity of the instrument was determined from the pertinence, relevance and clarity of each of the items, by experts (Pérez et al., 2015), In the case of this research, since the dimensions were found from the literature review, we proceeded to perform the content validity by means of a binomial distribution test using panels of experts, although it would have been more appropriate to raise, from the beginning, the idea of conducting an exploratory factor analysis.

The reliability of the instrument was based on the determination of Cronbach's Alpha, and included all the ordinal qualitative variables with their associated dimensions.

The statistic provided a result of 0.791, considered an adequate internal consistency value (Rodriguez and Reguant, 2020).

Central tendency and dispersion metrics for ordinal qualitative variables

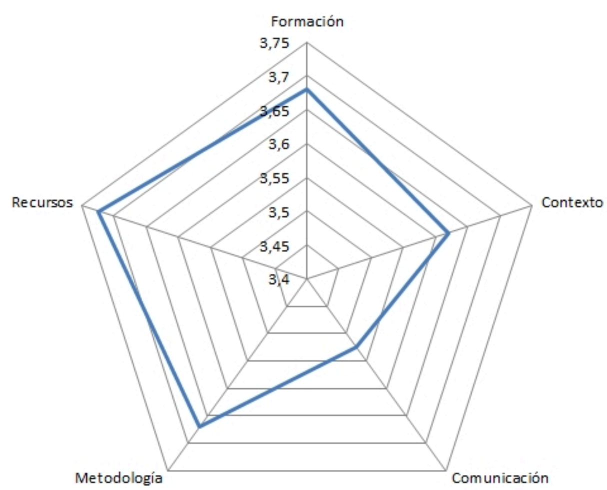
It could be observed that the mean overall satisfaction was 2.66/4 and that of the Age Group was in the range of 30-39 years. Likewise, the average entry profile was maintained between the Undergraduate/Diploma/Bachelor's Degree and the Graduate Degree.

In relation to the variable "Graduate Satisfaction", the item means ranged from 2.46 (item 5) to 2.79 (item 3).

The diagram in Figure 4 shows that the resources (delivery of materials and management and support of the virtual campus) were the most highly valued, in general, while communication and the possibilities of improving the social and economic context of the students were the least satisfactory aspects.

Figure 4

Radial diagram of the variable "Graduate Satisfaction"



Variable "Level of satisfaction"

Table 8 shows the descriptive data of central tendency and dispersion of this new variable.

Table 8

Basic statistics of the variable "Level of satisfaction"

	N	Minimum	Maximum	Media	Desv. Type
Satisfaction Level	186	25	43	33.7	5.53

Normality test

In order to determine the normal behavior of the variable and given that the sample was larger than 50 students, the Kolmogorov-Smirnov test was performed (De la Garza, Morales and González, 2013). The result is shown in Table 9.

Table 9

Kolmogorov-Smirnov normality test for the variable "Level of Satisfaction"

Normality tests						
Variable	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistician	**gl**	**Sig.**	**Statistician**	**gl**	**Sig.**
Level of satisfaction	0.193	186	0.000	0.810	186	0.000

Note: a. Lilliefors correction.

Since the resulting p-value (0.000) is less than 0.05, the null hypothesis of normal distribution of the data was rejected, so these values do not follow a normal distribution.

Categorization

The decision thresholds for the variable "Level of satisfaction" were determined on the basis of the following cut-off points (mean= 33.7, standard deviation= 5.53):

$$33.7 - 0.75*5.53= 29.55 \sim 29$$
$$33.7 + 0.75*5.53= 37.84 \sim 38$$

Thus, the data were grouped as shown in Table 10.

Table 10

Satisfaction level of graduates by grouping ranges

Level of satisfaction	Range	Frequency	%
Under	Values<=29	40	21.5
Medium	Values 30-38	112	60.2
High	Values>=39	34	18.3

Table 10 shows that 60.2% of the graduates have a medium level of satisfaction with the institution's *online* graduate programs, 21.5% have a low level of satisfaction and the rest feel highly satisfied after completing the corresponding graduate program.

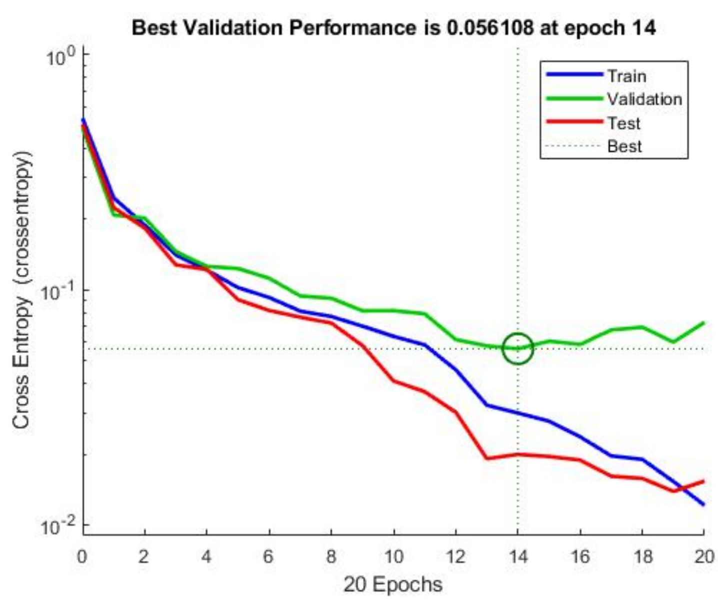
Neural network model performance

The performance of the model training, validation and testing data set is illustrated in Figure 5.

Cross entropy is preferred for classification, while mean square error is preferred for regression or predictions.

Figure 5

Cross entropy losses for training, validation and testing.



It was observed that training was terminated when the validation error did not improve during the six iterations following the 14thepoch, where the best training and validation results were obtained, with a minimum cross-entropy error value of 0.056108.

Also, after the 14th epoch, it is found that training follows a linear downward trend, while the validation loss function stabilizes in early stages, signifying some overfitting of the model to the training data. However, this overadjustment is not significant because the error is small.

Since the data to be obtained in the future are likely to be very similar to those used in the training set, the model can be considered valid; otherwise, other formulas should be sought to reduce overfitting.

Global confusion matrix

Once the training, validation and testing stages were completed, the global confusion matrix was obtained, which was used to analyze the sensitivity and accuracy of the model, among other characteristics (Table 11).

Table 11
Global confusion matrix

Output Class	*1*	**33**	**1**	**0**	97.1%
		17.7%	0.5%	0.0%	2.9%
	2	**1**	**111**	**4**	95.7%
		0.5%	59.7%	2.2%	4.3%
	3	**0**	**0**	**36**	100%
		0.0%	0.0%	19.4%	0.0%
		97.1%	*99.1%*	*90.0%*	*96.8%*
		2.9%	*0.9%*	*10.0%*	*3.2%*
		1	*2*	*3*	
			Target Class		

In the analysis by columns, it can be seen that, of the 34 graduates with a high level of satisfaction, 33 were classified correctly and only 1 incorrectly. Therefore, the false positive rate in this case was 2.9%. Similarly, of the 112 graduates with a medium level of satisfaction, 111 were classified correctly and 1 incorrectly, with a false positive rate of 0.9%. Finally, of the 40 graduates with a low level of satisfaction, 36 were classified correctly and 4 incorrectly, with a false positive rate of 10%.

In the analysis by ranks, it is observed that out of 34 projects identified as having a high level of sustainability, 33 were actually of high sustainability, which represents 97.1%, while the remaining 2.9% of the projects were actually of medium sustainability. Likewise, of the 116 projects identified as medium sustainability, 111 projects were actually medium sustainability, while 1 was high sustainability and 4 were low sustainability. Finally, of the 36 projects identified as low sustainability, all of them turned out to be really at this level.

It is observed that the overall accuracy is 96.8%, which indicates that the classification performed by the model is very good.

Example of testing

Table 12 shows the fit of some of the test data. It can be seen that what is predicted in a classification are probabilities, unlike regression, where an exact value is provided. For example, in the first case, the probability that a graduate with the characteristics shown has a medium level of satisfaction is 98.23%.

Table 12

Predicted probability values of the model in the test run

Graduate	Measurement criteria													Predicted values			Ranking
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**High**	**Medium**	**Under**	
1	3	4	3	3	3	4	3	3	3	2	3	3	3	.0004	.9823	.0173	Medium
2	2	4	3	4	4	4	4	4	3	3	3	4	4	.0266	.9715	.0019	Medium
3	4	4	3	2	4	4	4	2	2	1	2	2	4	.0000	.0025	.9975	Under
4	4	4	4	4	4	4	4	4	4	4	3	4	3	.9775	.0225	.0000	High
5 ..	4	4	4	4	3	3	3	3	3	2	3	2	3	.0024	.9955	.0021	Medium

Note: Source: Own elaboration, 2022

Discussion

This article investigated the need to develop a model to assess the satisfaction of graduates of various *online* graduate programs in the field of health and nutrition, based on a series of variables and dimensions, as interdependent parts of the whole.

It is noted in the literature the existence of a large amount of documented information at different stages of graduate training, initiatives and other models and tools, aimed at the same objective (Gonzalez et al., 2016). However, the lack of a single criterion when approaching research means that there are different approaches, some more complex than others.

In order to constitute the abstract reference, the literature review resulted in a model composed of nominal and ordinal qualitative variables, which included socio-cultural and demographic aspects of the graduate (age, gender, origin, program studied and entry profile), as well as another specifically referring to their satisfaction with online graduate programs, including four dimensions: methodology, organization, academic expectations and teaching work. This means that it is not an exhaustive model, since other types of variables (administrative, auxiliary services, etc.) were not taken into account. In this sense, there are virtual or non-virtual models that refer only to student satisfaction with university tutoring (Pérez et al., 2015). Other more complex research incorporates a greater number of dimensions and indicators; for example, at the margin of the teaching plan, teaching methodology and training, support services, administrative, enabling environment, level of self-realization, and infrastructure and facilities, among others (Fainholc, 2004; Álvarez et al., 2013; Mejías and Martínez, 2013; Pérez et al., 2015). It can be concluded that the lack of homogeneity of models focused on both virtual and face-to-face teaching means that there are different approaches, some more complex than others, and that, therefore, measuring the level of satisfaction of graduates turns out to be a complex task.

Obtaining a Likert scale measurement instrument, based on the degree of abstraction of the variables, is another point of disparity between models. This is because most of the bibliographic references consulted obtain the measurement criteria of the measuring instrument from indicators already contemplated in previous research by one or more authors. For example, this is the case of Álvarez et al. (2013), which are supported by the studies of Gento and Vivas (2003). In this research, however, the model variable "Graduate Satisfaction" was subjected to a process called operationalization, which underlies all the models instrumented, although its development is not made explicit in the literature reviewed (Reguant and Martinez, 2014). It is therefore necessary to obtain a measurement instrument that is as concrete as possible, without ambiguities, that allows the level of satisfaction of graduates to be measured directly by Pérez et al. (2015), rather than the difference between final perception and initial expectations (Mejías and Martínez, 2013).

In this research, when determining the reliability of the instrument, the results reflected a value of 0.791. This means that the instrument is consistent and provides good reliability (Rodriguez and Reguant, 2020). In this sense, all the models reviewed base the reliability of the instrument on the determination of the internal consistency coefficient "Cronbach's Alpha", applied to a test with a small number of students (Álvarez et al., 2013). The observed values in all cases are above 0.8 (Romo et al., 2012; Surdez et al., 2018), implying good reliability (Rodriguez and Reguant, 2020). It can be interpreted, therefore, as a high level of stability of the instrument due to the fact that there are not too many differences in results in its application to different realities.

The average overall satisfaction indicator was 2.66/4. This demonstrates a good level of participant satisfaction with the institution's *online* postgraduate programs in the field of health. It is also shown that about 80% of the participants presented a medium to high satisfaction with the health postgraduate programs in general. In this sense, it was found that there were no significant differences between the means of the dimensions of the "Graduate Satisfaction" variable, so these were attributed to chance.

In order to determine the strengths and weaknesses of the training process, the results showed that the management of the virtual campus, technical support and logistics for the delivery of didactic materials were the criteria most highly valued by the graduates, while communication and the facilities adopted by the institution to improve the socioeconomic context of the student were the lowest rated. This means that the institution must improve the external and internal communication channels with the student, as well as strengthen the policy of scholarships and other study aids, respectively. However, comparisons with other studies are disparate, since the reality is very diverse. For example, in the research by Álvarez et al. (2013) the variables that scored best, i.e., those in which students were most satisfied, were "Teachers' training and teaching skills" and "Students' level of self-realization". On the other hand, the "Infrastructure" and "Administrative services" variables were the items with the lowest level of satisfaction. Regarding the model of Romo et al. (2012), mention is made of a highly significant relationship between educational quality and student satisfaction in certain aspects of training, faculty, curriculum design and administrative organization; however, no dependence of satisfaction on gender issues or the career pursued was found. Similarly, the results of the research by Surdez et al. (2018) showed that 53% of the students felt some degree of dissatisfaction -partial or total-, especially with regard to infrastructure and the state of facilities, furniture and equipment, which in turn had an impact on the teaching-learning process. However, satisfaction was found with regard to self-realization and respectful treatment of the student by the tutors. Nor were significant differences found between satisfaction and some dimensions such as gender, average years in college and school cycle. For their part, Kuo, Walker, Belland, and Schroder (2013) conducted a pretest with a set of 111 students in the United States to measure their satisfaction in an online course. The study concluded that satisfaction was conditioned by ICT skills and that there were differences between gender, academic level (undergraduate and graduate) and time spent. In view of these results, the predominance of face-to-face versus distance learning models is confirmed, many of which do not take into account the socioeconomic context of the student, unlike the model proposed in this research.

After training and validation of the neural network, a minimum cross-entropy error value of 0.056108 was obtained just at iteration number 14. Likewise, after the testing stage with unused data, an overall accuracy of 96.8% was obtained. This indicates that the classification and prediction performed by the model is excellent. In this regard, no studies of these characteristics applied to e-learning graduates have been found, however, it is important to highlight the way in which artificial intelligence can complement statistics for decision making.

Conclusions

In this research article, a methodology was developed to classify and predict the level of satisfaction of a group of graduates of graduate programs in the health field.

Despite the existence of some models and standards, the literature review revealed the existence of a gap between the degree of quality of e-learning training and the level of satisfaction of graduates, mostly caused by the heterogeneity of existing models and the insufficiency of a methodology to determine the level of satisfaction of students, in general, and graduates, in particular.

The development of an instrument based on the operationalization of the variables of the model contributed to answering the research question in the affirmative and to elaborating a classification that, together with the answers obtained from the questionnaire, formed part of a machine learning model based on a neural network, which made it possible to establish forecasts with a high level of accuracy of 96.8% on the level of satisfaction of graduates for decision making. In this sense, more than 80% of the graduates presented a level of satisfaction between medium and high; however, the importance of strengthening the Institution's commitment to the social and economic circumstances of the student, facilitating access to possible scholarships or making the schedules of certain scheduled tasks more flexible, for example, was also highlighted.

Therefore, and in response to one of the research sub-questions, it is clear that artificial intelligence plays a fundamental role in reducing the gap between educational quality and the level of graduate satisfaction, as it can complement traditional statistics and interact with the most critical aspects that affect this context.

Recommendations

Finally, some recommendations are as follows:

- Perform a chi-square analysis to determine if there are relationships between variables, for example, to evaluate the degree of influence that variables such as origin, age groups, etc., have on the variable "level of graduate satisfaction", taking into account the non-normality of the data distribution.
- To verify the validity of the measurement instrument with an exploratory factorial design.
- Perform a comparison with other *online,* face-to-face postgraduate or undergraduate degrees (Pérez et al., 2015).
- Complement the results obtained with the opinions of the teaching staff (Llorent and Corbano, 2019).
- Expand the number of the sample with more participants, especially from the Eurasian area.

References

Álvarez, J., Chaparro, E.M., Reyes, D.E. (2015). Estudio de la satisfacción de los estudiantes con los servicios educativos brindados por instituciones de educación superior del Valle de Toluca. *REICE. Revista iberoamericana sobre calidad, eficacia y cambio en educación*, *13*(2). https://revistas.uam.es/reice/article/view/2788

Azcona, M., Manzini, F. (2019). La unidad de análisis y la unidad de observación. https://docer.com.ar/doc/v881vc

Brydges, R., Manzone, J., Shanks, D., Hatala, R., Hamstra, S. J., Zendejas, B., Cook, D. A. (2015). Self-regulated learning in simulation-based training: A systematic review and meta-analysis. *Medical Education*, *49*(4), 368-378. https://doi.org/10.1111/medu.12649

Castillo, J., Solórzano, B., Moreno, J. (2018). Design of a neural network for the prediction of the coefficient of primary losses in turbulent flow regime. https://machinelearningmastery.com/learning-curves-for-diagnosing-machine-learning-model-performance

De la Garza, J., Morales, B.N., Gonzalez, B.A. (2013). *Análisis Estadístico Multivariante*. McGraw Hill.

Domínguez, I. (2021). *Diseño y validación de una metodología para mejorar la experiencia de usuario de egresados latinoamericanos de maestrías en salud cursadas bajo metodología eLearning en un entorno Moodle adaptado.* [Unpublished doctoral dissertation]. Universidad Internacional Iberoamericana (UNINI-MX).

Faggella, D. (2018). *What is artificial intelligence? An informed definition. Emerj Artificial Intelligence Research*. https://emerj.com/ai-glossary-terms/what-is-artificial-intelligence-an-informed-definition.

Fainholc, B. (2004). La calidad en la educación a distancia continúa siendo un tema muy complejo. *Revista de Educación a Distancia (RED)*, *12.* https://revistas.um.es/red/article/view/25311

Fainholc, B. (2016). Presente y futuro latinoamericano de la enseñanza y el aprendizaje en entornos virtuales referidos a educación universitaria. *Revista de Educación a Distancia (RED)*, *48.* https://revistas.um.es/red/article/view/253431

Gento, S., Vivas, M. (2003). THE SEUE: Un Instrumento para Conocer la Satisfacción de los Estudiantes Universitarios con su Educación. *Pedagogical Action*, *12 (2),* 16-27. https://dialnet.unirioja.es/descarga/articulo/2972060.pdf

González Alonso, J., Pazmiño Santacruz, M. (2015). Cálculo e interpretación del Alfa de Cronbach para el caso de validación de la consistencia interna de un cuestionario, con dos posibles escalas tipo Likert. *Revista Publicando*, *2(1),* 62-67. https://nbn-resolving.org/urn:nbn:de:0168-ssoar-423821

González Sánchez, R., Tinoco Zermeño, M., Torres Preciado, V. (2017). Análisis de la satisfacción de la experiencia universitaria de los egresados en 2015 de la Universidad de Colima. *Economic Paradigm,* *8*(2), 59-84. https://paradigmaeconomico.uaemex.mx/article/view/4803

Hernández R., Fernández, C., Baptista, P. (2003). *Metodología de la investigación*, (3ª. Ed.). McGraw-Hill. http://catarina.udlap.mx/u_dl_a/tales/documentos/lad/pinera_e_rd/capitulo3.pdf

Kirkpatrick, D. L. (1994). *Evaluating training programs: The four levels.* Berret-Koehler Publishers.

Kuo, Y.C., Walker, A.E., Belland, B.R., Schroder, K.E.E. (2013). A predictive study of student satisfaction in online education programs. *The International Review of Research in Open and Distributed Learning, 14*(1), 16-39. https://doi.org/10.19173/irrodl.v14i1.1338

Llorent, V., Cobano, V. (2019). Análisis crítico de las encuestas universitarias de satisfacción docente. *Revista de Educación, 385*, 91-117. doi:10.4438/1988-592X-RE-2019-385-41

Marcelo, C., Zapata, M. (2008). Cuestionario para la evaluación. "Evaluación de la calidad para programas completos de formación docente a través de estrategias de aprendizaje abierto y a distancia". Metodología de uso y descripción de indicadores. *RED: Journal of Distance Education*, *7.*

Marciniak, R. (2015). Propuesta metodológica para la aplicación del benchmarking internacional en la evaluación de la calidad de la educación superior virtual. *RUSC Universities and Knowledge Society Journal, 12*(3), 46-61. https://doi.org/10.7238/rusc.v12i3.2163

Martinsuo, M., Huemann, M. (2021). Reporting case studies for making an impact. *International Journal of Project Management,* *39*(8), 827-833. https://doi.org/10.1016/j.ijproman.2021.11.005

McArdle, G. E. (1999). *Training Design and Delivery*. American Society for Training and Development.

Mejías, A., Martínez, D. (2013). Desarrollo de un Instrumento para Medir la Satisfacción Estudiantil en Educación Superior. *University Teaching, 10*(2). http://saber.ucv.ve/ojs/index.php/rev_docu/article/view/3704

Organización Mundial de la Salud (OMS). (2005). 58ª Asamblea Mundial de la Salud [Internet]. WHO. https://apps.who.int/iris/bitstream/handle/10665/23058/A58_2005_REC1-sp.pdf?sequence=1&isAllowed=y

Pedemonte, V. (2020). AI for sustainability: an overview of AI and the SDGs to contribute to the european policy-making. https://ec.europa.eu/futurium/en/system/files/ged/vincent-pedemonte_ai-for-sustainability_0.pdf

Pereira Campos S.A., Gelvez Pinto, L.N. (2018). Propuesta de un modelo latinoamericano para apoyar la gestión de calidad de la educación virtual. Un enfoque dinámico sistémico. In *IV Foro de Evaluación y Calidad*.

Pérez Cusó, F.J., Martínez Clares, P., Martínez Juárez, M. (2015). Satisfacción del estudiante universitario con la tutoría. Diseño y validación de un instrumento de medida. *Estudios sobre educación*, *29*, 81-101. https://doi.org/10.15581/004.29.81-101

Reguant Álvarez, M., Martínez Olmo, F. (2014). *Operacionalización de conceptos/variables*. Dipòsit Digital de la UB. http://diposit.ub.edu/dspace/bitstream/2445/57883/1/Indicadores-Repositorio.pdf

Rodríguez, J., Reguant, M. (2020). Calcular la fiabilidad de un cuestionario o escala mediante el SPSS: el coeficiente alfa de Cronbach. *REIRE Revista d'Innovació i Recerca en Educació, 13*(2), 1-13. https://doi.org/10.1344/reire2020.13.230048

Romo, J. R., Mendoza, G., Flores, G. (2012). *Relaciones conceptuales entre calidad educativa y satisfacción estudiantil, evaluadas con ecuaciones estructurales: El caso de la facultad de filosofía y letras de la Universidad Autónoma de Chihuahua.* http://cie.uach.mx/cd/docs/area_04/a4p11.pdf

Salinas, A. (2007). *Satisfacción del estudiante y calidad universitaria: un análisis explicatorio en la Unidad Académica Multidisciplinaria. Agronomía y Ciencias de la Universidad Autónoma de Tamaulipas*. [Doctoral dissertation]. University of Seville, Seville.

Shaw, T., Barnet, S., McGregor, D., Avery, J. (2015). Using the Knowledge, Process, Practice (KPP) model for driving the design and development of online postgraduate medical education. *Medical Teacher*, *37*(1), 53-58. https://doi.org/10.3109/0142159X.2014.92356

Shiferaw, F., Zolfo, M. (2012). The role of information communication technology (ICT) towards universal health coverage: The first steps of a telemedicine project in Ethiopia. Global Health Action, *5*(1), 1-8 https://doi.org/10.3402/gha.v5i0.15638

Schmidhuber, J. (2015). "Deep learning in neural networks: An overview. *Neural Networks, 61*, 85-117.

Surdez, E. G., Sandoval, M del C., Lamoyi, C.L., Sandoval, M del C., Lamoyi, C.L. (2018). Satisfacción estudiantil en la valoración de la calidad educativa universitaria. *Educación y Educadores, 21*(1), 9-26. https://doi.org/10.5294/edu.2018.21.1.1.

Torres, M., Karim, P. (2021). Size of a sample for a market research. Faculty of Engineering. Rafael Landívar University. *Electronic Newsletter, 2.* https://docplayer.es/424351-Tamano-de-una-muestra-para-unainvestigacion-de-mercado.html

UNE 66181:2012. Gestión de la calidad. Calidad de la formación virtual Madrid: AENOR.

Vann Slyke, C., Kittner, M., Belanger, F. (1998). Identifying Candidates for Distance education: A telecommuting perspective. In *Proceedings of the America's Conference on Information Systems*. Baltimore,

World Health Organization (WHO). (2018). Digital Health. Seventy-first World Health Assembly - Agenda item 12.4 (A71/A/CONF./1). WHO. http://apps.who.int/gb/e/e_wha71.html

Date received: 29/11/2022
Revision date: 11/12/2022
Date of acceptance: 22/12/2022