Virtual Education: Impact of Socio-Emotional and Pedagogical Factors on Academic Performance Based on Neural Networks and Stepwise Regression

José Luis Morales Rocha, Mario Aurelio Coyla Zela,
Nakaday Irazema Vargas Torres, Jarol Teófilo Ramos Rojas,
Santos Octavio Morillos Valderrama and Genciana Serruto Medina

August 13, 2021

Virtual Education: Impact of Socio-emotional and Pedagogical Factors on Academic Performance based on Neural Networks and Stepwise Regression

José Luis Morales Rocha, Doctor[1], Mario Aurelio Coyla Zela, Doctor[2], Nakaday Irazema Vargas Torres, Magíster Scientia [3], Jarol Teófilo Ramos Rojas, Doctor[4], Santos Octavio Morillos Valderrama, Doctor[5], and Genciana Serruto Medina, Doctor[6]

[1]Universidad Nacional de Moquegua, Moquegua, Perú, j*moralesr@unam.edu.pe*

[2]*Universidad Nacional de Moquegua, Moquegua, Perú, mcoylaz @unam.edu.pe*

[3]*Universidad Nacional de Moquegua, Moquegua, Perú, nvargast @unam.edu.pe*

[4]*Universidad José Carlos Mariátegui, Moquegua, Perú, jramos@ujcm.edu.pe*

[5]*Universidad Nacional del Altiplano, Puno, Perú, somorillos@unap.edu.pe*

[2]*Universidad Nacional de Moquegua, Moquegua, Perú, gserrutom@unam.edu.pe*

***Abstract*— The objective was to determine the impact of socio-emotional and pedagogical factors on the academic performance of students from the UNAM professional school of Management and Social Development during the COVID 19 pandemic, based on neural networks and stepwise regression. From the 38 indicators of socio-emotional and pedagogical factors applied to 236 students, the stepwise regression model identified 7 significant indicators (pedagogical and personal problems) that impact on academic performance, the neural network model determined the importance of indicators that influence academic performance. academic performance, considering the personal problems first, followed by pedagogical and family ones. The mean square error for the stepwise regression model is 7.25 and the coefficient of determination was 0.221, for the neural network model the mean square error was 6.78, while the coefficient of determination was 0.256. According to the results of both methodologies, the students of the professional school of Management and Social Development showed pedagogical and personal problems in their academic performance during the COVID 19 pandemic.**

Keywords— neural network, stepwise regression, COVID 19, pandemic, academic performance.

I. Introduction

The management of academic follow-up and monitoring of the teaching-learning process highlights its importance in the implementation of tutoring actions in order to know what are the factors that affect academic performance, with implications on performance of training and professional practice. Through this analysis process, it is also determined that in a critical period of pandemic by Covid19, the socio-emotional and family factors, have also affected and determined in a great deal the impact of the results of the summative and formative evaluations of the students considered in the sample and object study [1] [2].

Using the survey technique and with a digital questionnaire as an instrument, 38 indicators were collected within personal difficulties, pedagogical and family problems that influence the academic performance of students from the professional school of Management and Social Development of the National University of Moquegua. The stepwise regression analysis selected these as major indicators that influence academic performance of the students: failure of midterms, study techniques, difficulties to attend on time, difficulties joining the group, conflicts with fellow students, health and fitness problems and coexistence problems with a couple. Using the neural network model, the following indicators: difficulties joining the group, coexistence problems with a couple, difficulties due to a disability, failure of midterms, health and fitness problems, relationship with a teacher, feeding difficulties, problems due to beliefs, spirituality, religiosity, living alone were determined, followed at some distance by other predictors, as the most important indicators that influence the academic performance of students. The R2 of the regression model was 0.221 and its RSME was 7.248, of the neural network model the R2 was 0.256 and its RSME was 6.871.

The purpose of the research was to determine the impact of social-emotional and pedagogical factors on the academic performance of students from the school of Management and Social Development of the UNAM.

II. Related Works

This section presents the references of different investigations related to neural networks, stepwise regression and socio-emotional and pedagogical factors.

In [3] a predictive model for academic performance using neural network techniques on a data set of 300 students is described, the record was provided by a virtual learning environment developed in Moodle. A multilayer perceptron neural network was trained using a reverse propagation algorithm to predict the ability to successfully pass the studies program/ the race. The accuracy rate of the classification was high, averaging 75.28% for all subjects; concluding the effectiveness of the predictors obtained for the academic performance prediction.

In [4] it used an Artificial Neural Network model (ANN) and a Multiple Linear Regression model (MLR) to model the academic performance of university students. The accuracy of the models was rated using model evaluation criteria such as R2, NIC, MSE and ANIC. The modeling capacity of the ANN architecture developed was compared with an MLR model using the same training data sets. The squared regression prediction coefficients for the MLR and ANN models were 0.746 and 0.893, respectively. The results revealed that the ANN model compared to the MLR model, was more accurate in modeling the data set. This was because the ANN model had its MSE (0.072) compared to the traditional model, which MSE was 0.182.

In [1] techniques of artificial neural networks (ANN) with supervised learning, of the Multilayer Perceptron type (MLP) with the backpropagation algorithm (BP) were applied, to obtain a model capable of classifying factors that affect the academic performance of students, a comprehensive teachers evaluation, the publication of research products, and teachers who publish in high impact, regional journals, book chapters, from the databases of the Academic Control Computer System (SICOA) and the System of Publications of the Research Directorate (SPDI) of the National University of Chimborazo; the model generated useful knowledge for the decision-making process of the university community managers. As methodology, it used the Standard Industrial Hybrid Process for Data Mining (CRISP-DM). The reliability of the model was 94.72% and obtained factors, such as the academic performance of the students is excellent when the gender is female, single, not foreign, do not work, do not have children, but siblings, practicing sports and a cultural activity.

In [5] they implemented classification algorithms through the Python programming language such as decision tree, K the nearest neighbors, perceptron and others, which are compared to know the best prediction result. The gender and ICFES (Colombian state test) score for mathematical condition are in the upper range of all feature selection methods, and the perceptron algorithm provides better accuracy regarding the other algorithms. It concludes stating that the variables that influence the academic performance of engineering students are: age, gender, ICFES score for mathematical aptitude, ICFES global score, enrollment value and ICFES score for mathematical condition and cohort.

In the field of education, there are studies regarding the impact of socio-emotional and pedagogical factors on academic performance. In [6] they intend to identify “the main factors involved in the learning process, the information sources sufficiency and relevance, and the interest in the use of digital pedagogical tools for the learning process in orthopedic and traumatology surgical processes and protocols for surgical technicians. The academic performance of higher education university students in the city of Bogotá is affected by factors that impact poor academic performance: the institutional, pedagogical, psychosocial and socio-demographic ones. It was determined that the factors are the resources, tools and student motivation, planning and time organization, which permits to conclude that digital teaching resources which respond to the consolidation of concepts, helping to acquire procedural skills and the strengthening of attitudes and values of the student of surgical technology should be established, to achieve a significant learning process while their performance.

In [2] they determine the influence of emotional maturity on academic performance between the first and second year of medical studies of the students at the University of Sultan Zainal Abidin (UniSZa). The conclusion states that the intelligence and emotional maturity influence the academic performance of students, developing their enriching skills, it is shown that there is an association between emotional intelligence and the academic performance; furthermore, the studies include the development of awareness regarding the intelligence and the emotional maturity and its applications which represents a significant advantage for medical students.

As reported [7] by those who seek to find out the correlation between psychological well-being and learning styles, methodologies, social skills, and emotional intelligence of university students, have shown that the selected psychological and cognitive dimensions are positively correlated with the dimensions of psychological well-being, while it can be seen that the collaborative methodologies are more widely accepted.

In [8] an attempt to formulate an empirical model that explains the relationships between unobservable variables and observed variables, which influence the academic performance of high school students is made. The conclusion obtained is that the performance of college students corresponding to the period 2007-2010 was deficient in comparison with the national and the central-southern region of Mexico averages during the same period. It was observed that from 164 variables it was narrowed down to 57 which explain the influence on academic performance.

In [9] the relationship between emotional intelligence, test anxiety and academic stress among university students is investigated. It is inferred from the research findings, that female students have higher emotional intelligence than males; on the other hand, students do not differ significantly in emotional intelligence in relation to the course of study. We can see that low emotional intelligence is positively correlated with low academic performance. Regarding the variable test anxiety, there were no differences concerning the gender, but it can be seen that the students of medical sciences exposed greater anxiety before the exams than the psychology students. It is also appreciated that women can express their feelings, whereas men have the ability to control their impulses and tolerate stress. Therefore, it can be stated that emotional intelligence, test anxiety and stress is significant and predictive of university students’ academic achievement.

In [10] the purpose was to determine the students' problems in the academic, family, financial, social, personal, emotional and spiritual dimensions in order to evaluate the impact of these problems on their academic performance, to design student programs based on their needs. The research shows that students experience challenging problems in life and in their academic formation, these problems reveal the classroom pedagogy, the teachers’ failures and the lack of activities that strengthen the integral development of students as well as effective learning habits.

As [11] shows, the objective of the work was to find the relation between academic performance and behavioral regulation indices by means of the executive functions inventory BRIEF-A. It was found that the academic performance of university students is similar to the scales and indices of the BRIEF-A questionnaire, related to cognitive processing, behavior regulation, attention and emotional processing, which is important because it takes into account the influence of emotions in social contexts and its relation with academic performance in higher-level students, which shows the presence of behavioral indicators in the academic performance.

In [12] it points out that defining academic performance in higher-level students presents similar difficulties for its conceptualization and study than in other educational levels. However, in [5] it states that it is necessary to use variables of other factors that influence the academic performance such as university academic management, technological, Library, Institutional, pedagogical and intellectual factors. In addition, in [13] it is pointed out that the use of these digital spaces generates a trail of data, which linked to the students' grades, their behavior and their previous performance, as well as their assessment of adaptation to the various learning methods.

The usages of various models and methods of measurement and validation reveal that establishing the inconvenient of students during their learning process, affects their academic performance. Therefore, the aforementioned investigations mostly agree that the intervention in their learning processes would be the most coherent, according to the socio-emotional factors that challenge their academic training, during the time that we have been living. Virtual environments are increasingly trustworthy as they are being incorporated as part of the methodologies used to achieve convenient learning, improving their performance and behavior in previous, during and later stages.

III. THEORETICAL BACKGROUND

In order to differentiate neural networks from traditional statistical methods, the traditional linear regression model can acquire knowledge through the least square's method and store that knowledge in the regression coefficients, in this sense it becomes a neural network. In fact, it can be said that linear regression is a special case of certain neural networks. However, linear regression has a rigid model structure and a set of assumptions that are supposed before learning from the data. In [14] multiple regression methods and artificial neural networks are techniques used in many applications.

A. *Regression analysis*

In [15] regression analysis aims to determine the type of functional relationship that exists between a dependent variable (y) and one or more independent variables ($x_1, x_2, \ldots x_K$).

The multiple linear regression model is as follows:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k + \varepsilon \qquad (1)$$

where y is the answer and $x_1, \ldots, x_k$ are the regressors. The term linear is used because the equation is a linear function of the unknown parameters $\beta_0, \beta_1, \beta_2, \ldots, \beta_k$. The parameters $\beta_j = j = 0,1, \ldots, k$. are called regression coefficients. The parameter β_j represents the expected change in the response and per unit change in x_j when all other regressor variables $x_i (i \neq j)$ remain constant.

B. *Hypothesis test in multiple linear regression*

Once the model parameters have been estimated, two questions immediately arise:

1. What is the general adequacy of the model?

2. Which specific regressors seem to be important?

There are several hypothesis testing procedures that prove to be useful in answering these questions. Formal tests require that the random errors be independent and have a normal distribution with mean $E(\varepsilon_i) = 0$ and a variance $Var(\varepsilon_i) = \sigma^2$.

C. *Test of the significance of the regression*

The regression significance test is used to determine if there is a linear relationship between the response y and any of the regressor variables $x_1, x_2, \ldots, x_k$.

This procedure is usually considered as a general or global test of the adequacy of the model. The hypotheses are:

$$H_0: \beta_0 = \beta_1 = \ldots = \beta_k = 0$$

$$H_1: \beta_j \neq 0 \quad \text{at least for one } j$$

The rejection of the null hypothesis implies that at least one of the regressors $x_1, x_2, \ldots, x_k$ contributes significantly to the model. The test procedure is performed by analysis of variance, where the total sum of squares SS_T is divided into a sum of squares due to regression (SS_R) and a sum of squares of residuals (SS_{Res}).

The test procedure is usually summarized in an analysis of variance table.

TABLE I
ANALYSIS OF VARIANCE

Model	**Sum of Squares**	**df**	**Mean Square**	**F**
Regression	$\boldsymbol{SS_R}$	K	$\boldsymbol{MS_R}$	$\boldsymbol{MS_R/M_{Re}}$
Residual	$\boldsymbol{SS_{Res}}$	n-k-1	$\boldsymbol{MS_{Res}}$	
Total	$\boldsymbol{SS_T}$	n-1		

D. *Test on individual regression coefficients*

Once it has been determined that at least one of the regressors is important, the logical question is which one (s) is (are) useful? If a variable is added to a regression model, the regression sum of squares increases, and the residual sum of squares decreases.

It must be decided whether the increment of the sum of squares of the regression is sufficient to warrant the use of the additional regressor in the model. Adding a regressor also increases the variance of the fitted value $\hat{y}$, so care must be taken to include only regressors that have value to explain the answer.

Furthermore, if an unimportant regressor is added, the mean square of the residuals can be increased, thereby the utility of the model is minimized.

The hypotheses to test the significance of any individual regression coefficient, such as β_j, are:

$$H_0: \beta_j = 0$$

$$H_1: \beta_j \neq 0$$

If $H_0: \beta_j = 0$ is not rejected, it means that the regressor x_j can be eliminated from the model. The test statistic for this hypothesis is

$$t_0 = \frac{\hat{\beta}_j}{\sqrt{\hat{\sigma}^2 c_{jj}}} = \frac{\hat{\beta}_j}{se\ (\hat{\beta}_j)} \tag{2}$$

where c_{jj} is the diagonal element of $(X'X)^{-1}$ that corresponds to $\hat{\beta}_j$. The null hypothesis $H_0: \beta_j = 0$ is rejected if $|t_0| > \ t_{\alpha/2, n-k-1}$.

E. R^2 and R^2 adjusted

Two other ways to evaluate the general adequacy of the model are the R^2 and adjusted R^2 statistics; the latter is represented by R^2_{Adj}.

In general, R^2 always increases when a regressor is added to the model, regardless of the value of the contribution of that variable. Consequently, it is difficult to measure whether an increase in R^2 actually states something important. Some people who work with regression models prefer to use the adjusted R^2 statistic, which is defined:

$$R^2_{Adj} = 1 - \frac{SS_{Res}/(n-p)}{SS_T/(n-1)} \tag{3}$$

F. Stepwise regression

In [17] step regression is a mathematical approach to fitting regression models in which the introduction of candidate independent variables in the regression model is carried out one by one in a dynamic way. After each step in which a variable is added to the regression model, all existing variables in the model are checked to see if their significance has fallen below a pre-established tolerance level. If this is the case, this so-called non-significant variable will be removed from the regression model. This process will be carried out iteratively until no independent variable is included in the regression equation nor any independent variable is excluded from the regression equation, so that an optimal final set of independent variables can be established with the corresponding coefficients of adjustment

In [18] stepwise regression combines the mechanisms of inclusion and elimination of the independent variables by means of the forward and backward selection procedures.

G. Neural networks

The first study on neural networks dates back to 1943 when McCulloch and Pitts [19] created a simple neural network computer model. In 1949, Hebb introduced learning rules for neural networks. Since then, neural networks have grown rapidly and have been widely applied in various fields.

The term neural network is applied to a family of models related in an approximate way that is characterized by a great parameter space and a flexible structure and that comes from the studies on the functioning of the brain, [20] motivated by the structure of a biological neuronal system capable of processing in parallel like a brain.

A neural network is a massively parallel distributed processor with a natural propensity to store experimental knowledge and make it available for its use. It resembles the brain function in two ways:

1) *Knowledge is acquired through the network through a learning process.*

2) *Interneuronal connecting forces, known as synaptic weights, are used to store knowledge.*

H. Structure of the neural network

Although neural networks propose minimal demands on the assumptions and structure of the model, it is useful to understand the general architecture of the network. The multilayer perceptron network (MLP) is a function of predictors (also called inputs or independent variables) that minimizes the prediction error of the target variables (also called outputs)

This structure is called feedforward architecture because the network connections flow one-dimensionally [21] from the input to the output layer without feedback loops.

In the figure:

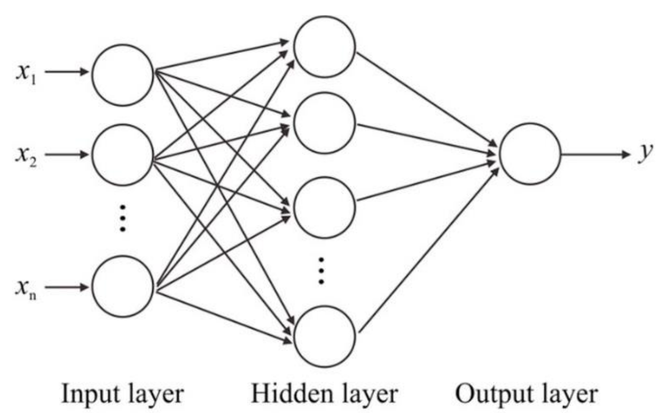

Fig. 1. Red neuronal

In [22], [23] A neural network normally consists of three layers: the input, hidden and output layers:

1) *The input layer* contains the predictors.

2) *The hidden layer* contains unobservable nodes (or units). The value of each hidden unit is a function of the predictors; the exact form of the function depends, on the one hand, on the type

of network and, on the other hand, on user-controllable specifications.

3) *The output layer* contains the answers. Each output unit is a function of the hidden inputs. One more time, the exact form of the function depends first on the type of network and also on user-controllable specifications.

Mathematically, a k neuron can be defined by the following two equations:

$$y_k = f(u_k + b_k) \quad (5)$$

$$u_k = \sum_{i=1}^{N} w_{ki}\, x_i \quad (5)$$

where x_1, x_2, x_3, ..., x_n denote the input signals, w_{k1}, w_{k2}, w_{k3},..., w_{kn} are the connection weights of the neuron, u_k is the linear output of the linear combination between weighted inputs, b_k is the bias term, f is the activation function and y_k is the neuron output signal

The MLP neural networks are trained on the basis of the backpropagation algorithm (BP) which follows a learning procedure based on the error correction rule. In fact, the network produces network outputs by processing the received input data. Then, by comparing the target values and the network output, the error value is calculated. Subsequently, the weights and biases are adjusted to minimize the error; the training process continues until the network reaches a predefined minimum permissible error.

I. Multilayer Perceptron

In [20] the multilayer Perceptron neural network (MLP) generates a predictive model for one or more dependent (target) variables based on the values of the predictor variables. In [24] supervised learning problems are applied and trained on a set of input-output pairs and learn to model the relationship between those inputs and outputs .

The MLP network allows a hidden second layer; in that case, each unit in the second hidden layer is a function of the units in the first hidden layer, and each answer is a function of the units in the second hidden layer.

J. Partitions

1) *The training sample* comprises the data records used to train the neural network; a certain percentage of cases in the data set must be assigned to the training sample in order to get a model.

2) *The test sample* is an independent set of data records that is used to track errors during training, in order to avoid over-training.

K. Accuracy of the proposed models

During the research, a comparative study was carried out to compare the performance and precision of the proposed MLP and Stepwise Regression models against socio-emotional indicators that influence academic performance.

The statistics models used in this study include the mean square error (RMSE) and coefficient of determination (R^2). These parameters can be calculated as follows:

$$RSME = \frac{1}{n}\sum_{i=1}^{n}(y - \hat{y})^2 \quad (6)$$

$$R^2 = \frac{\sum_{i=1}^{n}(\hat{y} - \bar{y})^2}{\sum_{i=1}^{n}(y - \bar{y})^2} \quad (7)$$

IV. METHODOLOGY

A sample of 236 students from the UNAM professional school of Public Management and Social Development was considered. Two models were used: stepwise regression and the Multilayer Perceptron neural network.

Three factors that determine the students' academic performance have been considered: problems that hinder academic progress, personal difficulties they face and family problems that students of the professional school of Management and Social Development of the National University of Moquegua go through.

The pedagogical problems that hinder the academic progress of students are determined by 14 indicators:

- Difficulties to attend on time, failure of midterms, difficulties working in a group, difficulties to make a presentation, difficulties doing work, relationship with a teacher, study techniques, study habits, vocation and identification with their major, study materials and supplies, problems due to the remote education methodology, low grades on assignments and tasks, computer availability for personal use and internet availability and connectivity for class sessions

The personal difficulties that students face are determined by 15 indicators:

- Health and fitness problems, difficulties due to a disability, feeding problems, housing problems, difficulties joining the group, conflicts with fellow students, conflicts with other close personal relationships, insecurity and fears, permanent stress, emotional problems, limitations in setting personal goals and aspirations, problems due to beliefs, spirituality, religion, little autonomy and decision making, students feel discriminated and feeling isolated and alone.

Family problems that students go through are determined by 9 indicators:

- Conflict in their relationship with a family member, living alone affects them, having suffered the loss of a close relative, having a sick family member, having family members who are financially dependent on the student, coexistence problems as a couple, overload in responsibilities of the paternal or maternal role, economic problems and conflictive coexistence at home

The variables of age and gender of the students have also been considered for the selection of variables. The variable academic performance has been determined by the weighted average of the students of the professional school of Management and Social Development grades.

V. RESULT

This section shows the results obtained from applying neural networks and stepwise regression.

A. *Selection of variables*

Multiple linear regression analysis was used, using the Stepwise method for the selection of independent variables that are the most significant ones.

TABLE II
MODEL SUMMARY[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
7	,471[a]	,221	,197	2,739223

a. Predictors: (Constant), Failure of midterms, Study techniques, Difficulties to attend on time, Difficulties joining the group, Conflicts with fellow students, Health and fitness problems, Coexistence problems as a couple
b. Dependent Variable: Academic performance

In Table II, a determination coefficient of 0,221 is observed, which indicates that the academic performance of students from the professional school of Management and Social Development is explained by 22.1% of personal difficulties, academic problems and family problems.

TABLE III
ANOVA[a]

Model	Sum of Squares	df	Mean Square	F	Sig.
Regression	486,387	7	69,484	9,260	,000[b]
Residual	1710,762	228	7,503		
Total	2197,149	235			

a. Dependent Variable: Academic performance
b. Predictors: (Constant), Failure of midterms, Study techniques, Difficulties to attend on time, Difficulties joining the group, Conflicts with fellow students, Health and fitness problems, Coexistence problems as a couple

Table III shows the Analysis of Variance (ANOVA) that offers information about the adequacy of the regression model to estimate the values of the dependent variable. Through the Snedecor F statistic, it is observed that the Sig. (P-value = 0,000) is less than 0,05 of significance, this means that the regression model is significant.

TABLE IV
COEFFICIENTS[A]

Model	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Std. Error	Beta		
(Constant)	13,107	,273		48,055	,000
Failure of midterms	-1,756	,432	-,246	-4,067	,000
Study techniques	1,292	,389	,198	3,322	,001
Difficulties to attend on time	-1,115	,417	-,162	-2,675	,008
Difficulties joining the group	-1,357	,485	-,170	-2,798	,006
Conflicts with fellow students	1,530	,666	,136	2,298	,022
Health and fitness problems	-1,363	,575	-,140	-2,371	,019
Coexistence problems as a couple	-3,358	1,610	-,123	-2,085	,038

a. Dependent Variable: Academic performance

Table IV shows the coefficients of the regression model that are significant, this means that these indicators are the ones that influence the academic performance of the students of the Professional School of Management and Social Development

The multiple regression model is given by:

$$\hat{y} = 13{,}10 - 1{,}75x_1 + 1{,}29x_2 - 1{,}11x_3 - 1{,}35x_4 + 1{,}53x_6 - 1{,}36x_6 - 3{,}35x_7 \quad (8)$$

Where:

$\hat{y}_i$: Academic performance

x_1 : Failure of midterms

x_2 : Study techniques

x_3 : Difficulties to attend on time

x_4 : Difficulties joining the group

x_5 : Conflicts with fellow students

x_6 : Health and fitness problems

x_7 : Coexistence problems as a couple

TABLE V
SIGNIFICANT FACTORS AND INDICATORS

Factors	Indicators	Sig.
Pedagogical problems	Failure of midterms	,000
	Study techniques	,001
	Difficulties to attend on time	,008
Personal difficulties	Difficulties joining the group	,006
	Conflicts with fellow students	,022
	Health and fitness problems	,019
Family problems	Coexistence problems as a couple	,038

In Table V it is observed that the factors with greater significance determined by the stepwise regression are of a pedagogical nature related to study and learning techniques due to the lack of personal discipline and negative stress regarding the receipt of results of summative evaluations and then they are the socio-emotional factors related to the social integration of the group, as well as to the conflicts in the interaction of the execution of teamwork, also health issues typical of Covid19 and others of a personal and family kind.

B. *Modeling using the Multilayer Perceptron*

The following are the results obtained through the use of a Multilayer Perceptron, which allows to determine the most important indicators so they can be considered in the analyzes that we present in the research.

TABLE VI
CASE PROCESSING SUMMARY

		N	Percent
Sample	Training	168	71,2%
	Testing	68	28,8%
Valid		236	100,0%
Excluded		0	
Total		236	

In Table VI, the results of the Multilayer Perceptron processing are shown. The neural network considers 71.2% of the data for training and 28.8% for testing, no excluded data are presented in the model.

TABLE VII
NETWORK INFORMATION

Input Layer	Covariates	38
	Number of Units[a]	38
	Rescaling Method for Covariates	Standardized
Hidden Layer(s)	Number of Hidden Layers	1
	Number of Units in Hidden Layer 1[a]	3
	Activation Function	Hyperbolic tangent
Output Layer	Dependent Variables	Academic performance
	Number of Units	1
	Rescaling Method for Scale Dependents	Standardized
	Activation Function	Identity
	Error Function	Sum of Squares

a. Excluding the bias unit

Table VII presents the information of the neural network, the input layer has 38 covariates, for the change of scale in the covariates the standardized method was used, the neural network has a hidden layer with three units, the activation function considered is the hyperbolic tangent. The output layer is made up of a dependent variable (Academic performance), the standardized method was used for the scale change, the activation function considered is the identity function and the mean square error was determined.

TABLE VIII
NEURAL NETWORK MODEL SUMMARY

Training	Sum of Squares Error	58,786
	Relative Error	,704
	Stopping Rule Used	1 consecutive step(s) with no decrease in errora
	Training Time	0:00:00.04
Testing	Sum of Squares Error	29,679
	Relative Error	,820

Dependent Variable: Academic performance
a. Error computations are based on the testing sample.

In Table VIII, the results of the learning and testing process of the neural network model are presented. The Multilayer Perceptron shows in its training process the Sum of Squares of the Error of 58.786 and a relative error of 0.704; on the other hand, in the test process of the model a Sum of Squares of the Error of 29.679 and a relative error of 0.820.

TABLE IX
PARAMETER ESTIMATES

Predictor		Predicted			
		Hidden Layer 1			Output Layer
		H(1:1)	H(1:2)	H(1:3)	AP
Input Layer	(Bias)	-,289	-,463	,270	
	A1	-,054	,675	,358	
	A2	-,574	,671	,779	
	A3	-,092	,078	-,325	
	A4	,597	,611	,264	
	A5	-,224	-,096	-,131	
	A6	,502	-,905	-,102	
	A7	,033	,024	-1,055	
	A8	,107	-,079	-,601	
	A9	-,061	-,172	-,343	
	A10	,316	,328	-,014	
	A11	,659	-,187	-,501	
	A12	-,569	,083	,559	
	A13	,479	,512	,530	
	A14	,629	-,338	,434	
	A15	,008	1,326	-,243	
	A16	-,421	,017	,403	
	A17	,687	-,627	-,166	
	A18	-,141	,315	,149	
	A19	-,473	1,332	,599	
	A20	,465	,308	-,340	
	A21	-,388	,090	-,468	
	A22	-,061	-,207	,420	
	A23	,107	-,003	-,533	
	A24	,448	-,006	,188	
	A25	,161	,490	,015	
	A26	-,086	-,276	,411	
	A27	-,370	,291	-,352	
	A28	-,039	-,423	,225	
	A29	,192	,120	-,471	
	A30	-,659	-,157	-,220	
	A31	,285	,347	,560	
	A32	,113	-,517	,660	
	A33	-,320	-,351	,071	
	A34	,224	-,037	-,163	
	A35	,321	,346	,515	
	A36	-,152	-,233	,090	
	A37	,346	-,448	,234	
	A38	,034	,113	-,273	
Hidden Layer 1	(Bias)				-,087
	H(1:1)				,221
	H(1:2)				-,379
	H(1:3)				-,350

Table IX shows the parameters of the neural network model (Multilayer Perceptron). The thicker lines are the negative synaptic weights, while the thinner lines are the synaptic weights with positive values.

TABLE X
INDEPENDENT VARIABLE IMPORTANCE

Independent Variables	Importance	Normalized Importance
Difficulties joining the group	0.078	100.00%
Coexistence problems as a couple	0.066	85.00%
Difficulties due to a disability	0.061	78.40%
Failure of midterms	0.052	67.00%
Health and fitness problems	0.05	64.40%
Relationship with a teacher	0.044	56.20%
Feeding problems	0.042	53.30%
Problems due to beliefs, spirituality, religiosity	0.037	47.60%
Living only affects you	0.036	46.60%

Table X shows the importance of the independent variables, this measure indicates how much the value predicted by the network changes for different values of the independent variable, the variables are ordered from greatest to least importance. Being the following variables difficulties joining the group, coexistence problems as a couple, difficulties due to a disability, failure of midterms, health and fitness problems, relationship with a teacher, feeding problems, problems due to beliefs, spirituality, religion, living alone affects them, difficulties to attend on time, etc. the most important used by the network to predict the academic performance.

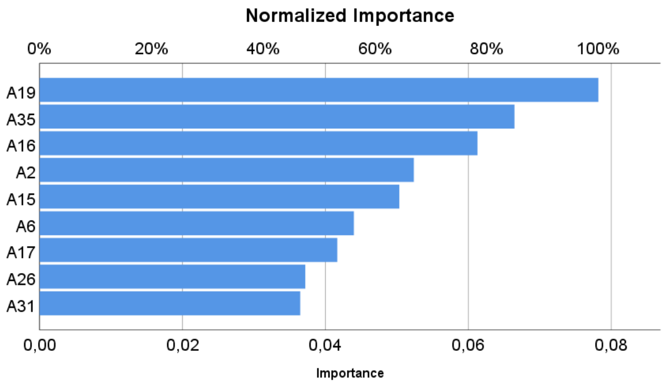

Fig. 2. Normalized Importance

In Fig. 2, the importance graph shows that the results of the academic performance of the students are dominated by difficulties joining the group, coexistence problems as a couple, difficulties due to a disability, failure of midterms, health and fitness problems, relationship with a teacher, feeding problems, problems due to beliefs, spirituality, religion, living alone affects them, followed at some distance by other predictors.

TABLE XI
FACTORS OF IMPORTANCE

Factors	Importance	Normalized Importance
Personal difficulties	0.454	100.0%
Pedagogical problems	0.338	74.4%
Family problems	0.208	45.9%

Table XI shows that the most important factors with information from indicators grouped in a hierarchical order through neural networks that have the greatest impact on the academic performance of students are personal difficulties and pedagogical problems.

C. *Accuracy of the proposed models*

In this part of the article, the performance and accuracy of the MLP and Stepwise Regression models are compared.

TABLE XII
ACCURACY OF MODELS

Factors	RSME	R2
Stepwise regression	7,248992147	0.221372182
Perceptron Multilayer	6,871364038	0.256741404

Table XII shows the RSME and the R^2 for the Stepwise regression and the MLP, showing better results with the Multilayer Perceptron.

VI. DISCUSSION

According to the objective, to determine the impact of socio-emotional and pedagogical factors on the academic performance of students of the professional school of Management and Social Development during the COVID 19 pandemic, based on neural networks and stepwise regression, the results of the model of Regression of Table V shows that failure of midterms, study techniques, difficulties to attend on time, difficulties joining the group, conflicts with fellow students, health and fitness problems and coexistence problems as a couple present significant coefficients, indicating that they are determining factors that influence students' academic performance; the results of the neural network presented in Table X show that difficulties joining the group, coexistence problems as a couple, difficulties due to a disability, failure of midterms, health and fitness problems, relationship with a teacher, feeding problems, problems due to beliefs, spirituality, religion, living alone affects them, followed at some distance by other predictors, are the most important indicators that influence the academic performance. Both models show that personal difficulties, pedagogical problems and family problems hindered the academic progress of students in times of the COVID 19 pandemic, results that when compared with what was found in [12] who determine that there is a difference between students with higher performance compared to the ones with lower performance in terms of socio-emotional skills and coping strategies. [2] demonstrates that the existence of a statistically significant positive correlation between emotional maturity and academic performance, these results affirm that personal difficulties, pedagogical problems and family problems are factors that influence the academic performance of students in the pandemic of COVID 19 times.

Regarding the precision of the models used: Stepwise Regression and Multilayer Perceptron neural network model, the results of Table XII show that the neural network model presents better results compared to the regression model in determining the factors that hinder the academic progress in the students of the professional school of Management and Social Development during the COVID 19 pandemic. In [5] it is concluded that the Perceptron algorithm is the most accurate model of all the algorithms they have used in their research. These results also show that neural networks get better results.

VII. CONCLUSIONS

This research aimed to state that the effectiveness of the multiple regression model selected by the stepwise method is clearly shown by the coefficient of determination of 0.221, explained by the interconnection of the input variables represented by personal difficulties, academic and family problems with an intensity of interaction of 22.1 % with the output function of academic performance, with a significance lower than 0.05, which shows that the regression model is associated and that the predictive addition is significant.

The investigation has concluded that the most significant factors are determined by the failure of the midterms, study techniques, difficulties to attend on time and difficulties to join the group, likewise conflicts with fellow students, health and fitness problems and coexistence problems as a couple.

Moreover, the constructed neural network has been able to determine the interconnection behavior of input variables composed of 38 covariates in the input layer and it is determined that they have no direct connection with neurons in the environment demonstrated only by a hidden layer with three units and the output layer is made up of the variable academic performance. Consequently, the neural network was not only used as a prediction tool but also to quantitatively explain the interrelation between the input and output variables.

The neural network model allows to evaluate with greater accuracy the level of importance of the independent variables, ranking them of most relevant to least relevant, the same that are determined by the difficulties joining the group, coexistence problems as a couple, difficulties due to a disability, failure of midterms, health and fitness problems, relationship with a teacher, feeding problems, as the most important ones used by the network to predict academic performance.

VIII. FUTURE WORK

Future work will focus on proposing models to analyze educational digital resources that guarantee the acquisition of procedural skills and the strengthening of attitudes and values of the student, to achieve meaningful learning.

Presenting the research using the model of structural equations that show that the personal and academic characteristics of the student are determined by the academic performance.

REFERENCES

[1] I. Quizhpilema Lazo, "Implementation of a prediction model based on artificial neural networks for the classification of academic and research information from UNACH," Universidad Nacional de Chimborazo, 2020.

[2] V. Bhagat, Y. Izad, J. Jayaraj, R. Husain, K. Che Mat, and M. Moe Thwe Aung, "Emotional Maturity Among Medical Students and Its Impact on Their Academic Performance," *Tost*, vol. 4, no. 1, pp. 48–54, 2017, [Online]. Available: http://transectscience.org/.

[3] N. Salgado Reyes, J. Beltrán Morales, J. Guaña Moya, C. Escobar Teran, D. Nicolalde Rodriguez, and G. Chafla Altamirano, "Model to predict academic performance based on neural networks and learning analytics," *Iber. J. Inf. Syst. Technol.*, pp. 258–266, 2018, [Online]. Available: https://search.proquest.com/openview/5720c78f2e17a27355a8766fe81feb1a/1?pq-origsite=gscholar&cbl=1006393.

[4] O. C. Asogwa, N. M. Eze, C. M. Eze, C. I. Okonkwo, and K. C. Ojide, "On the Efficiency of Multiple Linear Regression over Artificial Neural Network Models," no. February, 2020, doi: 10.5923/j.statistics.20201001.01.

[5] L. E. Contreras, H. J. Fuentes, and J. I. Rodríguez, "Prediction of academic performance as an indicator of success / failure of engineering students, using machine learning," *Acad. Train.*, vol. 13, no. 5, pp. 233–246, 2020.

[6] A. D. P. Ferrucho Suárez and M. P. Cepeda, "Factors that intervene in the learning of orthopedics and traumatology in students of surgical assistance in an institution of higher education in the city of Bogota," *Educ. Medica*, no. xx, 2019, doi: 10.1016/j.edumed.2018.12.009.

[7] F. M. Morales-Rodríguez, I. Espigares-López, T. Brown, and J. M. Pérez-Mármol, "The relationship between psychological well-being and psychosocial factors in university students," *Int. J. Environ. Res. Public Health*, vol. 17, no. 13, pp. 1–21, 2020, doi: 10.3390/ijerph17134778.

[8] R. R. Carreto, F. G. Jaimes, F. Julián Ariza Hernández, F. S. Rosas, and O. Freddy Torreblanca Ignacio, "An empirical model to explain the academic performance of high school students," *Educ. Profiles*, vol. 36, no. 146, pp. 45–62, 2014, doi: 10.1016/S0185-2698(14)70127-8.

[9] G. Stankovska, D. Dimitrovski, and S. Angelkoska, "Emotional Intelligence, Test Anxiety and Academic Stress among University Students," vol. 16, pp. 157–164, 2018.

[10] J. D. Dagdag, H. G. Cuizon, and A. O. Bete, "College students' problems and their link to academic performance: Basis for needs-driven student programs," *J. Res. Policy Pract. Teach. Teach. Educ.*, vol. 9, no. 2, pp. 54–65, 2019, doi: 10.37134/jrpptte.vol9.no2.5.2019.

[11] O. Reynoso Orozco and T. E. Méndez-Luévano, "Is it possible to predict academic performance? The regulation of behavior as an indicator of academic performance in higher education students," *Dialogues Educ. Curr. issues Educ. Res.*, vol. 9, no. 16, 2018.

[12] N. Piovano, M. Solodovsky, and G. Pascuali, "Socio-emotional competencies and stress . How they relate with academic performance in higher education students," *Ciencias Humanas y Soc.*, pp. 69–80, 2020, [Online]. Available: https://repositorio.unimoron.edu.ar/handle/10.34073/218.

[13] M. C. De la Iglesia Villasol, "Analytics of digital uses and academic performance. A case study with university students," *REIRE Innov. Mag.*, vol. 13, no. 2, 2020, doi: 10.1344/reire2020.13.229267.

[14] L. Ji, X. Ma, X. Dang, W. Ji, Q. Song, and S. Liu, "'Median paralyzing dose' and 'multiple regression analysis', a new viewpoint to the research method of spinal cord injury," *Med. Hypotheses*, vol. 140, no. 256, 2020, doi: 10.1016/j.mehy.2020.109677.

[15] D. Zhao, Y. Wang, D. Liang, and M. Ivanov, "Performances of regression model and artificial neural network in monitoring welding quality based on power signal," *J. Mater. Res. Technol.*, vol. 9, no. 2, pp. 1231–1240, 2020, doi: 10.1016/j.jmrt.2019.11.050.

[16] S. Searle and M. Gruber, *Linear Models*, Second. New Jersey, 2017.

[17] B. Pang and Y. Yin, "Development of a dimensionless rod-bundle CHF correlation based on stepwise regression method, Part I: determination of the basic form with CHF data obtained with uniform axial heat flux profile," *Ann. Nucl. Energy*, vol. 143, p. 107454, 2020, doi: 10.1016/j.anucene.2020.107454.

[18] R. K. H. Galvão, M. C. U. de Araújo, and S. F. C. Soares, *Linear Regression Modeling: Variable Selection*, vol. 3, no. September 2019. 2020.

[19] A. Hashemi Fath, F. Madanifar, and M. Abbasi, "Implementation of multilayer perceptron (MLP) and radial basis function (RBF) neural networks to predict solution gas-oil ratio of crude oil systems," *Petroleum*, vol. 6, no. 1, pp. 80–91, 2020, doi: 10.1016/j.petlm.2018.12.002.

[20] L. C. P. Velasco, R. P. Serquiña, M. S. A. Abdul Zamad, B. F. Juanico, and J. C. Lomocso, "Week-ahead rainfall forecasting using multilayer perceptron neural network," *Procedia Comput. Sci.*, vol. 161, no. 3, pp. 386–397, 2019, doi: 10.1016/j.procs.2019.11.137.

[21] F. A. Shaqra, R. Duwairi, and M. Al-Ayyoub, "Recognizing emotion from speech based on age and gender using hierarchical models," *Procedia Comput. Sci.*, vol. 151, no. 2018, pp. 37–44, 2019, doi: 10.1016/j.procs.2019.04.009.

[22] R. Štohl and K. Stibor, "Predicting Safety Solutions via an Artificial Neural Network," *IFAC-PapersOnLine*, vol. 52, no. 27, pp. 490–495, 2019, doi: 10.1016/j.ifacol.2019.12.711.

[23] T. Goudjerkan and M. Jayabalan, "Predicting 30-day hospital readmission for diabetes patients using multilayer perceptron," *Int. J. Adv. Comput. Sci. Appl.*, vol. 10, no. 2, pp. 268–275, 2019, doi: 10.14569/ijacsa.2019.0100236.

[24] B. Cortés, R. Tapia Sánchez, and J. J. Flores, "Characterization of a polycrystalline photovoltaic cell using artificial neural networks," *Sol. Energy*, vol. 196, no. December 2019, pp. 157–167, 2020, doi: 10.1016/j.solener.2019.12.012.