Regression Model and Neural Network Applied to the Public Spending Execution

José Morales[1]

E. P. Gestión Pública y Desarrollo Social
Universidad Nacional de Moquegua
Moquegua, Perú

José Huanca[2]

E. P. Gestión Pública y Desarrollo Social
Universidad Nacional de Juliaca
Juliaca, Perú

Abstract—**Artificial Neural Networks are connectionist systems formed by numerous process units called neurons connected to each other, which adapt their structure through learning techniques to solve problems of function approximation and pattern classification. They process information that is supplied to them, either to obtain relationships between them and the objective function that is intended to be approximated, or by classifying these data into different categories. Regression analysis aims to determine the type of functional relationship that exists between a dependent variable and one or more independent variables. The purpose of the research is to use regression methods (multiple regression) and artificial neural networks (multilayer perceptron) to determine the influence of spending execution on the regional government's public budget. 95% of the variability of the budget of Moquegua region has been determined and explained by the three sectors (primary, secondary and tertiary) and 5% is determined by other factors outside the regional government budget. The determination coefficients $R^2 = 95.9\%$ in the regression model and $R^2 = 95.3\%$ in the neural network (multilayer perceptron). It has been demonstrated that Artificial neural networks and regression models have obtained very similar results, achieving good and good-fit models.**

Keywords—Regression; neural network; multilayer perceptron; institutional budget; public spending

I. INTRODUCTION

In [1], intelligent systems present various inductive or deductive models in the process of acquiring knowledge for decision-making. However, this knowledge is insufficient to effectively model reasoning that is complex and also requires further development. [2] Artificial neural networks are information processing systems inspired by the behavior of biological neural networks. [3] Regression models are tools used to estimate and forecast future trends in variables with historical data. In [4], it is considered that in recent years there has been a growing interest in the use of artificial neural networks to predict and forecast.

In almost all countries there is the problem of administration in public institutions, they were always questioned for their inefficiency and ineffectiveness in managing and achieving their objectives, to the extreme that citizens associate public administration as a synonym for mismanagement. Regional and local governments have problems in the ability to execute spending, which has the effect that they are forced to refund the money from their budgets to the Public Treasure, evidencing damage to the population.

After evaluating the prediction results with the application of multiple regression models and neural networks (Multilayer Perceptron) using information from the institutional budget of the Moquegua regional government and the expenses executed in the primary, secondary and tertiary sectors of 11 years, they were observed determination coefficients $R^2 = 95.9\%$ in the regression model and $R^2 = 95.3\%$ in the neural network (multilayer perceptron) showing that the artificial neural networks and the regression models obtained very similar results, achieving an adequate goodness of fit in the two models.

The mean squared error of the regression model is 17.30, while the mean squared error of the model of the regression model (Multilayer Perceptron) is 20.08 million new soles, very similar results that confirm the goodness of fit of both models.

The content of the article is organized as follows: Section II shows a brief review of the state of the art in relation to regression analysis and artificial neural networks; In Section III, some concepts and definitions are included that will allow a better understanding of the content of the work; In Section IV, the results obtained are described and analyzed; Section V describes the conclusions reached at the end of this investigation; finally it shows the future work that can be implemented to improve the results achieved.

II. RELATED WORKS

This section presents the references of different investigations related to artificial neural networks and regression techniques.

Linear multiple regression techniques and artificial neural networks are applied [3] to predict wear in the context of remaining height and remaining life of mineral mill linings.

They used artificial neural networks and support vector machines in the construction of prediction models for the S&P 500 stock index [5], the RBF kernel SVM model provided good prediction capabilities with respect to the artificial neural and regression models.

They analyze the effect of the number of neurons, the input data and the training function in neural models [6], in addition, they develop multiple regression models, analyzing the adjusted R squared to determine the input parameters that are

significant. In this way, they seek to develop an efficient alternative method to calculate natural frequencies in steel beams with various geometric characteristics in terms of boundary conditions.

In [7], they make predictions of house prices based on NJOP house prices in the city of Malang using regression models and particle swarm optimization (PSO). They demonstrated that combined regression and PSO are adequate methods to obtain minimum prediction errors (IDR 14,186). They performed several tests with regression models and particle swarm optimization.

They used two prediction methods: multiple regression and artificial neural networks (multilayer perceptron) with the objective to predict the surface roughness in AISI 316l dry steel turning [8]. For its implementation, cutting parameters such as speed, feed and machining time have been considered. The results demonstrated that the artificial neural network technique shows better precision than multiple regression.

In [9], they present and compare simple, multiple and multivariate linear regression models with the perceptron multilayer neural network model. The multilayer perceptron model is similar to a regression model, due to the similarity of the output variable that is related by applying an activation function through a linear combination of its weights (coefficients) with input variables (predictor variables). The results show that, for the simple linear regression case, an R^2 value of 96.8% for neural networks, higher than the regression model (87.5%). In multiple regression, both techniques show similar R^2 results (97.8% and 98.2%). In the multivariate regression, for the first dependent variable, the R^2 values for the neural network and the linear regression were 22.0% and 54.0%, while, in the second dependent variable, the values were 87.0% and 66.5%.

III. THEORETICAL BACKGROUND

Regression methods and artificial neural networks are techniques used to determine dependency relationships. In this article, two modeling techniques were used: multiple regression and multilayer perceptron in order to determine the influence of spending execution on the public budget.

A. Linear Regression

The objective of the regression analysis [8] is to determine the type of functional relationship that exists between a dependent variable (Y) and one or more independent variables $(x_1, x_2, ... x_K)$.

Regression analysis is one of the most widely used tools in estimating and predicting future trends in variables through the analysis of historical data [3]. It is an extremely flexible procedure that can aid decision-making in many areas, such as sales, expenses, weather forecasting, etc. Regression is a technique used to predict the value of a dependent variable using one or more independent variables. Mathematically:

$$Y = a + b_1x_1 + b_2x_2 + \cdots + b_kx_k \qquad (1)$$

Where Y is the response variable, *a* and b_i i = 1 ... k are regression coefficients, x_1, x_2 ...... x_k represent the explanatory variables, and K is the total number of explanatory variables.

The values of the parameters a, b_1, b_2, b_3, …, b_K are estimated using the least squares estimation to determine optimal values of the parameters of the regression model [5]. R-squared, the coefficient of determination is used as a performance measure to determine how successfully the method explains the variation in the data [10].

It is important to note that most regression analyzes are based on the use of a statistic called p-value or P-Value, which corresponds to the probability of accepting the null hypothesis, compared to the level of significance a (it was used α = 0.05) [10]. However, the purpose of performing a regression analysis is to determine how the independent variables are related to the response variable.

B. Artificial Neural Networks

Neural networks are tools that can solve classification and prediction problems [11]. Neural networks are self-adaptive methods based on methods that fit data without any explicit specification of functional form for an underlying model and can also approximate any function with a certain precision.

Artificial neural networks are information processing systems that are based on performance characteristics of biological neural networks. Artificial neural networks are developed as generalizations of mathematical models of neural biology or human cognition based on the following assumptions [2] [11] [12] [13]:

1) Information processing occurs at many simple nodes (nodes are also called cells, units, or neurons).

2) Feeder links are used to pass signals between nodes.

3) Each feeder link is associated with a weight, which is a number that multiplies the signals.

4) Each cell applies a trigger function (usually non-linear) to the weighted sum of its input to produce an output.

Fig. 1 shows the architecture of a multilayer perceptron neural network model.

Artificial neural networks are techniques that seek to build intelligent programs using models that simulate the network of neurons in the human brain [6]. Artificial neural networks are biologically inspired computer models consisting of various processing elements (neurons). Neurons are connected to elements through weights that make up the structure of neural networks. Artificial neural networks have elements to process information, such as transfer functions, weighted inputs and outputs. Artificial neural networks are made up of one layer or several layers of neurons [13] [14].
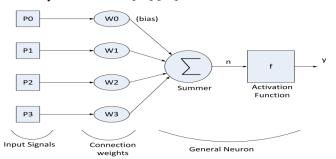

Fig. 1. Mathematical Model of a Basic Neural Network.

In [2], the inputs of the artificial neural network are modified by weights (synapses in the biological neural network). A positive weight represents an exciting connection and a negative weight means an inhibitory connection. A dummy entry, which is known as bias, is used in training. The weighted inputs are summed linearly. Finally, an activation function is applied to the weighted sum to determine the range and output characteristics of the artificial neural network.

A multilayer perceptron (MLP) is considered as an artificial neural network architecture widely used in predictive analysis [15]. [16] Multilayer perceptron neural network is an artificial neural network used for classification, pattern recognition, and prediction.

The multilayer perceptron (MLP) is the simplest form of artificial neural networks consisting of three layers. It starts with an input layer, has a hidden layer, and ends with an output layer. Each layer is made up of one or more neurons.

The multilayer perceptron collects information via the input layer, then processes it using an activation and weighted sum function, and the process ends at the output layer [17]. In order for the artificial neural network to achieve better performance and to deal with non-linear problems, the artificial neural network model is more complex.

In [18], multilayer perceptron neural network consists of one input layer, several hidden layers, and one output layer. The neurons of each layer are connected with the neurons in the following layer, that is, the input neuron is connected with the hidden layer neuron which connected with the output neuron, but the neurons of the same layer are not connected to each other.

The input layer's neurons receive the data then pass the data into the next layers until reaching the output layer.

IV. RESULTS

This section shows the results obtained from comparing the multiple regression model and with a multilayer perceptron (neural network) applied to the execution of public spending in the Moquegua regional government of 11 years. The data is given in millions of new soles.

A. Research Variables

The variables used are the following:

Y: BIM (Budget Institutional Modified)

X_1: Expenditure made in the primary sector

X_2: Expenditure made in the secondary sector

X_3: Expenditure made in the tertiary sector

The primary sector is made up of economic activities related to the extraction and transformation of natural resources into primary products.

The secondary sector is linked to craft and manufacturing industry activities.

Finally, the tertiary sector is the one dedicated to offering services to society and companies.

According to Table I, it is observed that the regional government of Moquegua was assigned an average budget of 494.90 million of new soles per year, in the primary sector an average of 48.29 million of new soles was executed, in the secondary sector 54.88 million of new soles and in the tertiary sector 274.08 million of new soles.

B. Regression Analysis

According to Table II, a determination coefficient of 0.959 is observed, which indicates that the productive sectors affect the regional government budget by 95.9%.

Table III shows the ANOVA that provides information on 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. Is less than 0.05, this means that the three sectors determine the execution of budget spending assigned to the regional government.

According to the results of Table IV, the multiple regression model is:

$$Y = 93,0111 + 0,326x_1 + 1,743x_2 + 1,060x_k \qquad (2)$$

TABLE I. DESCRIPTIVE STATISTICS

Variables	N	Mean	Std. Deviation
Budget Institutional Modified	11	494,906998	89,993105
Primary sector	11	48,293048	30,418761
Secondary sector	11	54,883252	17,709831
Third sector	11	274,081853	83,048184
Valid N	11		

TABLE II. MODEL SUMMARYB

R	R Square	Adjusted R Square	Std. Error of the Estimate
,979a	,959	,942	21,693842758

TABLE III. ANOVA

Model	Sum of Squares	df	Mean Square	F	Sig.
Regression	77693,231	3	25897,744	55,029	,000
Residual	3294,360	7	470,623		
Total	80987,590	10			

TABLE IV. COEFFICIENTS

Model	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Std. Error	Beta		
(Constant)	93,011	36,480		2,550	,038
Primary sector	,326	,261	,110	1,249	,252
Secondary sector	1,743	,462	,343	3,770	,007
Third sector	1,060	,086	,978	12,317	,000

The multiple regression model shows a coefficient for the primary sector of 0.326, for the secondary sector 1,743 and in the third sector 1,060 in millions of new soles. The regression model coefficients represent the mean changes in the response variable (BIM) for a unit of change in the predictor variable while keeping the other predictors in the model constant.

The scatter diagram shows the existence of a positive and linear relationship between the BIM and the prediction of the regression model. Fig. 2 shows the relationship between BIM and MLR prediction.

The predicted residuals graph shows a scatter plot of the residuals (the observed value minus the predicted value) on the Y axis, the predicted values on the X axis. The maximum positive value of the residuals is 19.96 and the minimum value is negative is -33.93 million of new soles. Fig. 3 shows the predicted residuals of the MLR.

C. Artificial Neural Networks

Table V presents the summary of the processing of cases

The neural network assigns as a training sample 90.9% of the cases and 9.1% to the test sample. The neural network does not present any case excluded in the analysis.
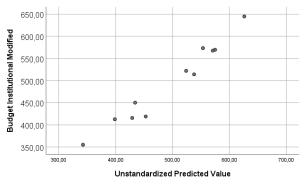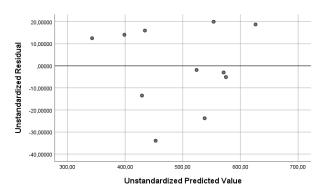

Fig. 2. Scatterplot between BIM and MLR Prediction.

Fig. 3. Predicted Residuals of the MLR.

TABLE V. CASE PROCESSING SUMMARY

		N	Percent
Sample	Training	10	90,9%
	Testing	1	9,1%
Valid		11	11
Total		11	

TABLE VI. NETWORK INFORMATION

Input Layer	Covariates	1	Primary sector
		2	Secondary sector
		3	Third sector
	Number of Units		3
	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	1	Budget Institutional Modified
	Number of Units		1
	Rescaling Method for Scale Dependents		Standardized
	Activation Function		Identity
	Error Function		Sum of Squares

Table VI shows the neural network information (Multilayer Perceptron) that is useful to ensure that the specifications are correct.

1) The input layer has three units, which is the number of independent variables (X_1, X_2, X_3).

2) The neural network has a hidden layer and the procedure has chosen three units in the hidden layer.

3) An output unit is created for the scale dependent variable (Y). Its scale is changed according to the typified method, which requires the use of the identity activation function for the output layer.

4) The normalization technique was used to change the scales of the covariates. The minimum is subtracted and divided by the range, $(x - min) / (max - min)$. Normalized values are between 0 and 1.

5) A sum of squares error is reported since the dependent variables are scale.

Table VII shows the neural network weights in the hidden layer and the output layer. The final weights of the neural network from the input layer to the hidden layer are: -0.111; -0.037; 0.163; -0.362, 0.307; 0.344; -1,064, 0.016; 0.023; The weights from the hidden layer to the output layer are: -1,174; -0.363; 0.394.

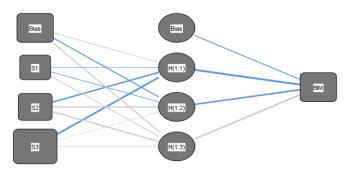Fig. 4 shows the Neural Network Model (Multilayer Perceptron).

The neural network model summary (Table VIII) displays information about neural network training and test results. The relative error during training is 0.059; at the test stage it is not determined because they are low for the neural network model.

A scatter plot is presented with the predicted values on the Y axis, the observed values on the X axis. Ideally, the values should be located along a 45-degree line starting at the origin. When examining the graph, it is observed that the neural network (multilayer perceptron) makes a good prognosis. Fig. 5 shows the relationship between BIM and MLP prediction.

TABLE VII.   PARAMETER ESTIMATES

Predictor		Predicted			
		Hidden Layer 1			Output Layer
		H(1:1)	H(1:2)	H(1:3)	BIM
Input Layer	(Bias)	,111	-,224	,269	
	S1	-,029	-,037	,163	
	S2	-,362	,307	,344	
	S3	-1,064	,016	,023	
Hidden Layer 1	(Bias)				-,174
	H(1:1)				-1,389
	H(1:2)				-,363
	H(1:3)				,394

Fig. 4.   Multilayer Perceptron Model

TABLE VIII.   MODEL SUMMARY

Training	Sum of Squares Error	,266
	Relative Error	,059
	Stopping Rule Used	1 consecutive step(s) with no decrease in error
	Training Time	0:00:00.00
Testing	Sum of Squares Error	,002
	Relative Error	.

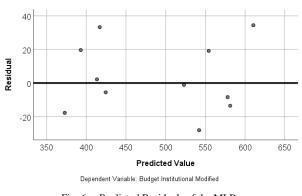Fig. 5.   Scatter Diagram between BIM and MLP.

Fig. 6.   Predicted Residuals of the MLP.

The predicted residuals graph shows a scatter plot of the residuals (the observed value minus the predicted value) on the Y axis, the predicted values on the X axis. The maximum positive value of the residuals is 34,394 and the minimum negative value is -28,099 million of new soles. Fig. 6 shows the predicted residuals of the MLP.

The importance of an independent variable is a measure that indicates how much the value predicted by the network model changes for different values of the independent variable. It seems that the variables related to the Modified Institutional Budget (BIM) of the regional government (the secondary sector and the tertiary sector) have the greatest effect on the PIM at 23% and 69% respectively. The primary sector has an effect of 8% in the budget assigned to the regional government. It could be said that the Moquegua regional government attaches greater importance to transportation, communications, environment, sanitation, housing and urban development, health, culture and sport, education, social protection and social welfare (tertiary sector). Fig. 7 shows the importance of the MLP variables.
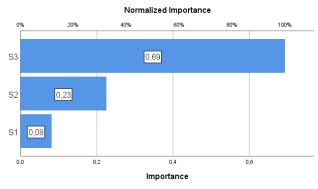

Fig. 7.   Importance of the MLP Variables.

V.   DISCUSSION

To validate the neural network and regression models, we calculated the coefficient of determination R^2 = 95.9% in the regression and R^2 = 95.3% in the neural network (multilayer perceptron). Therefore, we can affirm that 95% of the variability of the regional government budget is explained by the three sectors and 5% is determined by other factors outside the budget. In conclusion, we consider that artificial neural networks and regression models manage to obtain very similar results, achieving good good-fit models (Fig. 8).
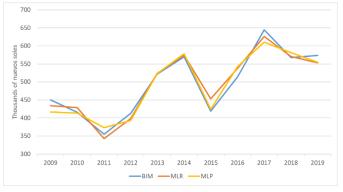
Fig. 8.   Regression Model - Neural Network Model.

In [19], neural networks are powerful tools used widely used for building prediction models [20]. In regression studies, as well as artificial neural network models, regression analysis has been shown to be one of the most widely used methodologies for expressing the dependence of a response variable on several independent variables [21]. Multiple Regression and Neural Network techniques can be effectively used to make predictions.

VI. Conclusions

This article has proposed and developed a multiple linear regression model with one dependent variable and three independent variables, a neural network (multilayer Perceptron) with an input layer that has three units, a hidden layer with three units, and an output unit for the dependent variable of type scale. Its scale is changed according to the standardized method, which requires the use of the identity activation function for the output layer. It has been determined that 95% of the variability of the budget of the Moquegua region is explained by three sectors (primary, secondary and tertiary) and 5% is determined by other factors unrelated to the regional government budget. The determination coefficients R^2 = 95.9% in the regression model and R^2 = 95.3% in the neural network (multilayer perceptron). The mean squared error of the regression model is 17.30, while the mean squared error of the neural network model (multilayer perceptron) is 20.08 million of new soles, very similar results that confirm the good goodness of fit of both models. Finally, it is concluded that the artificial neural networks and the regression models show very similar results for this case, achieving an adequate goodness of fit of both models.

In this work, research efforts have focused on certain specific questions, the application of regression techniques and artificial neural networks in models of public spending execution at the national level have been reserved for future work.

References

[1]   V. Yashchenko, "Multidimensional Neural-Like Growing Networks - A New Type of Neural Network," Int. J. Adv. Comput. Sci. Appl., vol. 6, no. 4, pp. 1–10, 2015.

[2]   M. K, I. A., and G. Onwodi, "Neural Network Based Hausa Language Speech Recognition," Int. J. Adv. Res. Artif. Intell., vol. 1, no. 2, pp. 39–44, 2012.

[3]   F. Ahmadzadeh and J. L. -, "Application of multi regressive linear model and neural network for wear prediction of grinding mill liners," Int. J. Adv. Comput. Sci. Appl., vol. 4, no. 5, pp. 53–58, 2013.

[4]   S. M. Cabaneros, J. K. Calautit, and B. R. Hughes, "A review of artificial neural network models for ambient air pollution prediction," Environ. Model. Softw., vol. 119, no. June 2018, pp. 285–304, 2019.

[5]   A. F., S. Elsir, and H. Faris, "A Comparison between Regression, Artificial Neural Networks and Support Vector Machines for Predicting Stock Market Index," Int. J. Adv. Res. Artif. Intell., vol. 4, no. 7, pp. 55–63, 2015.

[6]   M. Avcar and K. Saplioglu, "An Artificial Neural Network Application for Estimation of Natural Frequencies of Beams," Int. J. Adv. Comput. Sci. Appl., vol. 6, no. 6, pp. 94–102, 2015.

[7]   A. Nur, R. Ema, H. Taufiq, and W. Firdaus, "Modeling House Price Prediction using Regression Analysis and Particle Swarm Optimization Case Study : Malang, East Java, Indonesia," Int. J. Adv. Comput. Sci. Appl., vol. 8, no. 10, pp. 323–326, 2017.

[8]   Y. Morales Tamayo, Y. K. Zamora Hernández, P. J. Vásquez Carrera, M. P. Porras Vásconez, J. L. Bárzaga Quesada, and R. J. López Bustamante, "Network And Multiple Regression For Comparison Between Artificial Neural The Prediction of Superficial Roughness In Dry Turning," Ingenius, no. 19, pp. 79–88, 2018.

[9]   C. Menacho Ch, "Lineal regression models with neural networks," vol. 75, no. 2, pp. 253–260, 2014.

[10] C. M. Frau and R. M. Guajardo, "Comparison of neural networks and linear regression to estimate site productivity in forest plantations, using geomatic," Bosque, vol. 27, no. 1, pp. 35–43, 2006.

[11] A. Khalid and N. A., "Determining the Efficient Structure of Feed-Forward Neural Network to Classify Breast Cancer Dataset," Int. J. Adv. Comput. Sci. Appl., vol. 5, no. 12, pp. 87–90, 2014.

[12] N. M., A. Al, and S. A., "Load Balancing with Neural Network," Int. J. Adv. Comput. Sci. Appl., vol. 4, no. 10, pp. 138–145, 2013.

[13] A. Yudhana, I. Riadi, and F. Ridho, "DDoS classification using neural network and naïve bayes methods for network forensics," Int. J. Adv. Comput. Sci. Appl., vol. 9, no. 11, pp. 177–183, 2018.

[14] S. M. Snineh, M. Youssfi, A. Daaif, and O. Bouattane, "Micro agent and neural network based model for data error detection in a real time data stream," Int. J. Adv. Comput. Sci. Appl., vol. 10, no. 7, pp. 171–177, 2019.

[15] 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.

[16] B. Pandey, P. K. Bhanodia, A. Khamparia, and D. K. Pandey, "A comprehensive survey of edge prediction in social networks: Techniques, parameters and challenges," Expert Syst. Appl., vol. 124, pp. 164–181, 2019.

[17] 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.

[18] A. H. Elsheikh, S. W. Sharshir, M. Abd Elaziz, A. E. Kabeel, W. Guilan, and Z. Haiou, "Modeling of solar energy systems using artificial neural network: A comprehensive review," Sol. Energy, vol. 180, no. October 2018, pp. 622–639, 2019.

[19] M. Daoud and M. Mayo, "A survey of neural network-based cancer prediction models from microarray data," Artif. Intell. Med., vol. 97, no. October 2018, pp. 204–214, 2019.

[20] R. Kumar, R. K. Aggarwal, and J. D. Sharma, "Comparison of regression and artificial neural network models for estimation of global solar radiations," Renew. Sustain. Energy Rev., vol. 52, pp. 1294–1299, 2015.

[21] H. H. Mohamad, A. H. Ibrahim, and H. H. Massoud, "Assessment of the expected construction company's net profit using neural network and multiple regression models," Ain Shams Eng. J., vol. 4, no. 3, pp. 375–385, 2013.